

Eluna Learns AI Session Q & A

**What or is there a size limit on the csv file you can upload into code interpreter?**

Aaron Krebeck: None that I have found. I have uploaded csvs with over a million rows. Aaron Krebeck

**Did you try the Alma Enumerator tool before this new workflow and is this one more efficient?**

Aaron Krebeck: The Alma enumerator tool uses the item description as a starting point and then creates enum/chron entries. In our case we needed to clean and standardize the item descriptions first. As part of that process, the script basically did the work of the enumerator tool for us. So it wasn't a perfect match for our use case.

**How do you connect ChatGPT to Alma?**

Aaron Krebeck: We pulled the existing item description data out of Alma via Alma Analytics and then loaded the csv into ChatGPT to do the work. Then we downloaded the csv with the finished work and loaded the information back into Alma via a python script that can be found here: https://github.com/WRLC/alma-notes-import-flask.

**This sounds amazing and I want to do a project like this. Would you be willing to share your code?**

The code for adding the new item descriptions is based in Flask and Python and can be found here: https://github.com/WRLC/alma-notes-import-flask

**If projects like this are scaled up, there will come a time when the results are too large for humans to check for errors. So, at some point you might have item descriptions that are updated improperly, but that are not caught and corrected. What can be done to improve error catching, and what is an acceptable error rate?**

I believe we are already at that point. Our project actually involved error checking code that would parse the input and output and confirm that no numeric data was lost in the transformation. It's not impossible that bad data would slip through, but in our case, we are already working with flawed/out-of-spec data. So, the work will either improve it, do nothing, or it will continue to be flawed just in a slightly different way.

**Which API are you using (I am at Harvard)?**

Laura Morse: Do you mean specific call info? Or HOLLIS (Primo) APIs in general? More info on requesting access to HOLLIS (primo) APIs here. <https://wiki.harvard.edu/confluence/pages/viewpage.action?pageId=240453785>

**Is the end goal of this kind of project to have a full Primo integration akin to Bing/Copilot and Bard/Google integration where students/users can query their research questions with full sentences and get in-depth results to fit those questions from the LLM API?**

There are long-term aspirations to build much more comprehensive capabilities that may involve full-text semantic search with the ability to ask questions about research topics, beyond just the availability information provided currently.

**How did you reach the decision to start by updating the item description first and then update enumeration and chronology? It seems it might be useful to start by asking AI to parse the specifics first and then rebuild the description using description templates or additional details.**

In most cases, we didn't have anything in the enumeration and chronology field at all. So we worked with what we had.

**Does ChatGPT use the data you uploaded to become part of the training material available to all ChatGPT users? Is that fine with you?**

I don't know for sure, but in this case, we are fine with it. There is no user data or even possibly proprietary metadata like an OCLC number.

**Has anyone raised privacy concerns about inputting actual, unedited patron questions into ChatGPT since things that are put into Chat GPT become part of its knowledge base?**

Jonathan Young: This is definitely a concern that anyone who works with commercial AI systems should be thinking of. The OpenAI API does not save data or incorporate it into the training set, as opposed to the web ChatGPT version. Still, you should make sure that there is no personally identifiable educational data in your application.

**How much have you set aside in those AI budgets?**

Jonathan Young: Not much at the moment! At the moment everything is exploratory.

**Would you please put the url for the python script in the chat or share it after the session?**

Aaron Krebeck: <https://github.com/WRLC/alma-notes-import-flask>

**You mentioned that there is an Excel tool that can also achieve batch load changes of the Enum / Chron field. Where is that exactly?**

Aaron Krebeck: <https://github.com/wtee/alma_enumerator>

**With regards to item description uniformity, I'd like to hear more about how your decision tree. I was always told to record the information as it is on the publication. If the journal said "volume" you used v., if it said "Jahrgang" you used "Jahrg." -- but what if it used both in different years?**

Aaron Krebeck: It is the policy of our Metadata Committee to use the English version of the enumeration A field when possible. we definitely have some Jahrgang in our catalog and our work didn't change that to "volume", but we wouldn't touch item descriptions that had non-English enumeration.

**Would you be willing to create custom GPTs with your prompts?** [**https://chat.openai.com/gpts**](https://chat.openai.com/gpts)

Willing, but not currently able due to resource restraints. that said, the prompts we used were highly tailored to the unique ways that our item descriptions at each partner library were out of spec with our own consortial metadata guidelines. I’m sure a custom GPT would be a good starting point, but you’d really want to do your own.

**What sources have you found particularly helpful to learn how to prompt in that way (i.e., with a number of clauses shown to improve performance)?**

Jonathan Young: OpenAI provides some good advice in their documentation: <https://platform.openai.com/docs/guides/prompt-engineering>

You can also find information by searching on different prompt frameworks such as RTF, Chain of Thought, RISEN, etc.

**Have you considered using an Open Source model like Falcon or Llama 2?**

Jonathan Young: Currently this would require more expertise and investment than the commercial API models. But this would be a great path for libraries with more resources that want more control over their models and data.

**Since it has the library opening hour information, can it answer questions about the library hours?**

Katie Amaral: Great question. Since this is a limited pilot, currently it's only able to answer questions on what it's been prompted on specifically. We would need to build in a prompt for it to be able to accurately answer questions about hours, and that is one of the many capabilities we are planning to build at a later phase.

**Do you have future plans that use Chat GBI with Alma?**

Nothing concrete.

**What was the percentage that required manual examination after AI review?**

Zero. We had a separate automated script for error handling. If errors were detected, we did not make changes to that record. We were able to correct 80-90% in a fairly short period of time. It wouldn't be cost effective to do any amount of manual review.

**Can you ask it a higher-level question, like what call number range should I go to for books on [subject]?**

Katie Amaral: Currently it's only able to answer questions on what it's been prompted on specifically. Currently it is not prompted on call number ranges but that could be one of the many capabilities we are considering at a later phase.

**What happens if you submit offensive language?**

Katie Amaral: It would behave similarly to any general-purpose AI model, since that's what's being used in the backend. The demo shown in the presentation was a very limited scope pilot project, and in order for this to be production capable, we'll need to look at ways to build in filters and safeguards to avoid inappropriate context.

**With the ability to submit files, were you able to do the analysis in bulk? We've saved lots of emails but never had time to code the data. (Not READ scale, problem reports, but similar concept.)**

Jonathan Young: Yes, you can use the API to send text in bulk for analysis.

**Can Chat GPT be used to map other call numbers, particularly ones that are proprietary such as Dewey?**

Jonathan Young: Yes, you should be able to use the same procedure by providing any classification system in context in the prompt with the query.

**Were you surprised by anything you learned while working on this project?**

Two things, I was surprised at how well it worked, and I was surprised at ChatGPT's ability to create an error-checking script to check its own work and make sure we didn't end up with anything worse than what we started with.

**Can you please add these questions for the Harvard folks: Has this been shown to users and what was their response? Also, once a user clicks the links to Primo, won't the user just stay there to continue their searches? Do you think that might lead to search confusion once the user moves into the native Primo search?**

Katie Amaral: Testing sessions have been conducted by the User Research Center and questions were asked regarding potential use cases, viability of the tool, and what capabilities would be necessary for this to be a useful augmentation to standard HOLLIS search. Yes, certainly the tool would need many more capabilities and the UI would need to be built into HOLLIS in order to be suitable for Production, for example a chat window that appears on the bottom of the screen in HOLLIS (rather than an entirely separate window). The demo shown in the presentation was a limited scope pilot project, and the UI was not a main focus, so it's very minimalist and limited currently.

**Can you send the link to your github?**

Jonathan Young: <https://github.com/UHM-JonathanYoung/GPTChatRefAnalysis>