
EE/CprE/SE 491 BI-WEEKLY REPORT 3

02/11/24 – 02/24/24

***Group number:* 49**

***Project title:* Using Generative AI to Assess Learning**

***Client &/Advisor:* Dr. Henry Duwe**

Team Members/Role:
Alex Vongphandy - Backend and Prompt Developer
Akpobari Godpwer - Backend and Prompt Developer
Abram Demo - Frontend Developer
Drake Rippey - Frontend and Backend Developer

Weekly Summary*:* This week, on the backend portion of the project, the ReAct agent prompts to the Langroid framework, which is a more flexible and powerful tool for building LLM-powered applications. We tested GPT-4 on ReAct Agent/Langchain but found it too slow and unreliable for our purposes. We decided to switch to Langroid and converted our existing prompts into tasks and agents. We also customized some of the standard functions of Langroid, such as the output display, the variable passing, the task and agent functionality, and the chat history parsing. We designed a data structure to store information from the database and incorporated the rubric criteria into the question-generation prompts. We developed a grading prompt template and a student evaluation agent that will assess the student's responses based on the criteria. We demonstrated a prototype of our Langroid application to the client and received positive feedback. Some of the pending issues are to determine which agents are needed for the framework, and to develop a dynamic value system for each of the criteria. The frontend primary connections were created. This includes the ability for a student to click into a quiz from Canvas and have the frontend logic automatically send the student to the correct quiz and retrieve it from the backend. More parameters were added to the backend to make quiz storage easier. Frontend logic was also added which handles a launch request, determines what type of user sent the request, and sends them to either a quiz creation page or quiz page depending on if they are a teacher or not. There is also logic in place for if a quiz has not been generated yet.

Past week accomplishments

- Test GPT-4 on ReAct Agent/Langchain - Alex

    - Did not seem to function in a timely manner. In assessments, we need the responses to be concise and fast. Brought this to the attention of the group and client during the meeting.
- Research alternative LLM Framework - Alex
    - Langroid was chosen to replace Langchain due to its more modularity and functionality.
    - Familiarized and researched Langroid documentation and use
    - Prototyped a small demonstration of how the framework worked for weekly client meeting
- Langchain to Langroid conversion - Alex
    - Converted all current ReAct agent prompts into “tasks” and agents in Langroid
    - Able to limit follow-up questions properly using Pydantic models
- Langroid modular setup - Alex
    - Overridden several standard functions to meet our overall goal with the framework. This includes the display of output, passing of variables, task, and agent functionality. Chat history was redone and is now able to be parsed for AI to read.
- Data integration from the database into the backend - Akpobari
    - Designed a comprehensive data structure to store essential information, including critical class details, rubric specifications, and instructor data retrieved from the database.
    - Incorporated rubric criteria retrieved from the database into the question generation prompts such that each question is generated based on the criteria provided by the session_criteria variable.
- Developed Grading prompt template - Akpobari
    - In the prompt template, we will evaluate the student's quiz history with a response to the outlined criteria provided by the instructor. Prompt includes the evaluation of relevance of each response and assess whether the student's conversation is fully addressed in the responses
- Created Student Evaluation Agent - Akpobari
    - Defined prompt template for the student evaluation agent that will asses the relevance, completeness, and the depth of understanding created by the student. Originally created as a reACT agent so it will need to be migrated to the Langroid framework.
- Implemented Django Framework - Drake
    - After some research about frameworks that would make the building website process easier, I settled on Django. Using Django I set up a basic website and started making the pages needed for instructors, such as Quiz Creation. Django also allows for easy Database integration and OAuth Authentication.
- Django and Canvas Integration - Drake & Abram
    - Implemented a working connection between the canvas-based API and the Django website. This allows for all information needed to be pulled from canvas and used

in the various ways we need, such as quiz and rubric creation. We have yet to set up canvas post requests.

- HTTPS configuration - Abram
    - Canvas Requires an HTTPS connection for an lti to run. Using the Django framework and a generated OpenSSL certificate, we were able to run the program on HTTP and thus could get end-to-end connections between Canvas and the DB.
- OAuth handshake - Abram
    - Implemented the ability to be able to check all incoming connections using the lti launch URL to make sure that each one is from a valid OAuth provider and contains all of the correct information, including the OAuth signature and OAuth nonce.
- Grade Passback - Abram
    - Some work was done to try and pass grades back to canvas. This includes trying to masquerade as a student to submit an assignment on behalf of another student to work around the lti grade passback.

Pending issues

- Determine which agents are needed for the framework - Alex, Akpobari
- Develop a dynamic value system for each of the criteria to accommodate for the main question and the followup questions.
- Fully develop the website and “beautify” the pages. - Drake, Abram
- Implement a chainlit instance into the website. - Drake
- Handling CSFR requirements on HTTPS requests/getting HTTPS certificates. -Abram
- Grade Passback authorization and/or masquerading with the API. - Abram

Individual Contributions

Name	Individual Contributions	Hours (bi-weekly)	Hours cumulative
Alex Vongphandy	Langroid research and development, Langroid conversion, agent setup and functionality	14	34
Akpobari Godpower	Agent prompt engineering, Pydantic value prompt engineering, Data integration	12	30
Abram Demo	HTTPS configuration, frontend webpages, frontend logic	12	32
Drake Rippey	Database Remote Connection, Database Python Connection, Database Test Cases, Chainlit Redirecting/Styling	12	32

Plans for the upcoming week

- Research Langroid - Alex, Akpobari
    - Further research into documentation and full capabilities of Langroid. This is a new framework for us which has high-level documentation to base on the understanding of the inner workings of the code.
- Integrate mySQL functionality into Langroid - Alex, Akpobari

    - Be able to pull rubric into question generation.
- Migrate grading agent to langroid- Alex, Akpobari
    - Be able to grade the main question along with any follow up questions in a concise and fair manner.
- Develop a dynamic value system for each criterion to accommodate the main and the follow-up questions. - Alex, Akpobari
- Assessing the correct method for providing the grading agent the student responses. - Alex, Akpobari
- Allow our website to send post requests to canvas and grade student submissions - Drake & Abram
- Finish setting up all of the site's various pages including student quiz selection, and instructor view (general). - Drake
- Implement API masquerading as a fallback for grade passback. - Abram
    - If grade passback doesn't end up working, there is a workaround that requires getting an API key from an instructor and 'masquerading' as a student to submit a grade
- Research/Implement work to handle CSRF on HTTPS requests. -Abram
    - Although HTTPS is currently running, it's running in a dev environment and doesn't encrypt anything yet. In the future, we will need to be able to handle the CSRF information as well as have our own certificates.
- Continue working with grade passback authorization and submission format. - Abram

Broader Context addition

1. Have we identified or become aware of new effects?

While not new, one effect that was not mentioned in section 4.4 is the ethical concerns of using AI in regards to the fairness of grading. Since the grading is based on the interpretation of the student responses in comparison to the criteria, if the student effectively understands the criteria but has trouble communicating their understanding, it would be harder for the evaluation to recognize the student's understanding.

2. How can we argue for or provide evidence of positive effects?

We can advocate for the positive effects of utilizing AI-powered learning tools by individualizing the assessment experience for students and also providing them the platform to display an in-depth perspective of the understanding and application of the material. We can provide evidence by displaying the follow-up questions that, when answered correctly, can display a deeper illustration of what the student can comprehend, analyze, and create in relation to the main and follow-question generated by the AI

3. How can we address or justify negative effects?

One way to justify the negative effects of the ethics of AI on the fairness of grading is by being fully transparent to users. We will log all interactions between AI and the user. We can then use this information to develop better prompts or improve our model. Instructors will also be

provided a PDF transcript of the interaction such that if the student feels as if their responses were not interpreted correctly, they can approach the instructor with a valid form of the transcript.