02/24/24 – 03/29/24

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, we did some fine-tuning and adjustments to the "follow-up question generated" Langroid agent. We are still encountering issues with the AI "running on" or failing to terminate, which we will address this weekend before showing the client our fully integrated product. We mainly adjusted the prompts in an attempt to reduce the "running on" in this case, the agent does not select any tools, and the only way to terminate is if a tool is selected. The front end is getting close to being complete in terms of functionality. We are also working on the consistency of grading through updating the prompts. The front end now directs users to the correct place with the correct quiz. Meaning students get sent to a chainlit instance and teachers get sent to a quiz edit page. Oauth is also now completely working and handles grade passback and launch validation. We have now been spending time connecting everything up and making the frontend pages. In the following weeks, We will be deciding on a website style and making each page use this style to give the site an inviting feeling and make it look more professional.

Past week accomplishments

- Tested and diagnosed AI backend - Alex, Akpobari
  - Did end-testing on the AI to mimic the user taking a mock-quiz. Discovered some deficiencies that involve "run-on" functions and complexity with questions.
- Integrated Lecture Note Retrieval - Alex
  - Modified some parameters to try to get processing time lower.
  - Modified to only run during the creation of the main question.

- Prompt Engineering Testing - Alex, Akpobari
  - We have no official guide to prompt engineering, so testing different versions of prompts to help better understand the output of the agent tools.
- Prompt Engineering Testing - Alex, Akpobari
  - Working towards improving the prompt engineering for the grading agent.
- Database Model Revisions - Abram, Drake
  - We needed to modify and add objects to our database to suit the AI logic better.
  - Each quiz now only has one rubric, with multiple questions
  - The database now has a configuration object that holds a consumer key and consumer secret, which are used to verify launches
- Frontend Work - Abram, Drake
  - Pages added include an lti configuration page, rubric creation, and edit page.
  - the quiz detail page has been updated to show the current rubric, as well as some buttons to edit the current rubric
- OAuth Verification - Abram
  - Now, anytime someone tries to access the launch page, their post request is checked against the OAuth information, and if their signature isn't valid, they won't be allowed to continue.
- Merging frontend and backend - Drake
  - In previous weeks, we have had a very barebones version of the chatbot that didn't have any AI connectivity or any of the new langroid implementations.
  - A nearly complete backend with all up-to-date database connectivity is now available.

Pending issues

- AI failing to terminate/run on/waste of tokens. - Alex
- Ensure quiz is more robust and effective. - Alex, Akpobari
- Grade Passback issue where grading only happens for specific LTI configurations. - Abram
- Current Rubric in the database doesn't match up with how the AI is using it. - Abram
- Dockerize Langroid-based chatbot to allow for a full scale working prototype - Drake

Individual Contributions

Name	Individual Contributions	Hours (bi-weekly)	Hours cumulative
Alex Vongphandy	Diagnosing, Testing, Refining, planning mitigations for "run-on" AI. Hours are low due to being sick after Spring Break.	8	42
Akpobari Godpower	Refining grading agent prompt	7	41
Abram Demo	frontend webpages, frontend logic, Database model management, OAuth validation	12	42

Drake Rippey	Database Remote Connection, Database Python Connection, Database Test Cases, Chainlit Redirecting/Styling, Django implementation	13	45

Plans for the upcoming week

- Resolve issue with AI "run-on" - Alex
- Try to make quizzes more robust/effective. - Alex
- Prepare grading data for canvas - Akpobari
- Fix issue where grade passback only works for one configuration. - Abram
- Continue to update frontend webpages. -Abram
- Finish remaining db connections/Dockerize langroid - Drake
- Have a presentable working prototype - Drake

Midterm Feedback Section

There were a lot of questions about the AI behaviors, thresholds we set to model our desired behaviors, benchmarking criteria, fairness in grading and exam delivery (questions), and AI hallucinations. There was some additional interest in changing our presentation or delivery of our material, most notably defining and showing how Langroid is implemented into our design. The reviewing group praised us for our iterative approach, simple and clean design, and using the tools that are already available. One emphasis is to refine and fine-tune our design to ensure we are accomplishing our client's goals. One strength we received in our review was the fact that we chose an iterative approach in developing the agents that would be carrying out tasks in the backed. This approach allowed us to modularize our progress to ensure that components working within the system can work compatibly with each other. We also received feedback on the potential bias from the generated questions. Our solution to that was to incorporate potential questions generated in the instructor configuration so that the question generation agent is provided examples of what each question would potentially look like.