Names: Azaan Barlas, Kaiyao Ke, Chris Randall, Justin Huang

Project Progress Report

Tasks that have been completed:

- Working on language model using dataset we retrieved from all CS 410 lectures
- Completed data collection – transcripts of every lecture video (from weeks 1-12)
- Implemented a CNN model that classifies texts into groups. The model currently supports auto-parsing user feedback.
- Figured out existing pretrained model (OpenAI's GPT-J and BERT) to fine tune
- Figured out fine-tuning methodology (QA pairs)

Tasks that are pending:

- Out of the pretrained models we have found we need to choose which one to use. We're thinking about using a model from Open AI but still need to look into exactly which one we want to use.
- We might consider using our CNN classifier to assign a label (i.e. which week of content does the answer of the user's question belong to?) to user input text so that we narrow the scope of possible answers.
- Fine-tune our model and feed the data that we have collected from the lectures to the model.
- Use the model to create a chatbot experience so the user can ask questions about the class or the lectures.
- Test our chatbot to determine the accuracy of the responses given by the model.

Challenges:

- To fine tune our model well enough to answer questions on it similar to how chatgpt would answer on trained data
- Creating the chat bot