Weekly Progress Report

Group Members:

- Javier, Geron Simon
- Macapallag, Mhar Andrei
- Valdeabella, Seanrei Ethan

Work Completed:

This week, we made significant progress on the CodePath chatbot project. Our primary focus was on analyzing the errors our model was making. We utilized the Evaluation.ipynb notebook to run the model against our evaluation dataset and meticulously examined the misclassified examples. To gain a comprehensive understanding of the model's performance, we calculated various metrics, including accuracy, precision, recall, F1-score, BLEU score, and perplexity. We also generated visualizations such as the confusion matrix, ROC curve, response length distribution, and BLEU score distribution. This analysis revealed that the model struggled with complex or nuanced queries, occasionally generating repetitive or slightly off-topic responses. It also showed some sensitivity to minor variations in input phrasing.

Based on these findings, we adjusted our training process. We revisited the data preprocessing steps outlined in `Text_Preprocessing.ipynb`, ensuring its robustness and addressing potential edge cases. We also fine-tuned several hyperparameters within the `Chatbot Training and Function.ipynb` file, including learning rate, batch size, and the number of training epochs. After implementing these changes, we retrained the model. Concurrently, we began drafting the mid-term project report. We extracted relevant information from the provided research paper to articulate the project's goals, describe the dataset and its preparation, outline the model implementation, and document the training process. We also dedicated a section to the challenges encountered during the project.

Challenges Encountered:

Our main challenges this week revolved around refining the model and ensuring its accuracy. Fine-tuning the model's hyperparameters proved to be a complex and iterative process. It required a significant amount of experimentation to identify the optimal settings that balanced model complexity with training efficiency. Additionally, managing context in extended conversations presented difficulties. The chatbot sometimes failed to recall or utilize information from previous turns effectively. Another challenge was accurately detecting user intent, especially when confronted with ambiguous or multi-faceted queries. The model occasionally misinterpreted the user's goal or request, leading to inaccurate or irrelevant responses. Lastly, we encountered issues with the GPU memory when training and running the model which prevented us from working on the model.

Solutions Implemented:

To address the challenges, we implemented several solutions. We refined our data preprocessing pipeline, particularly the `preprocess_text()` and `correct_spelling()` functions in `Text_Preprocessing.ipynb`, to handle a broader range of input variations and improve data consistency. We conducted a series of experiments to optimize the hyperparameters, systematically adjusting them and evaluating the model's performance after each change. We explored techniques like grid search to efficiently navigate the hyperparameter space. To improve context management, we reviewed and adjusted the `max_history` parameter in the `AdvancedChatbotManager` class. We also experimented with strategies for summarizing or prioritizing past conversation turns. For intent detection, we refined the `detect_user_intent()` function, incorporating more nuanced natural language understanding techniques and expanding our list of intents and keywords. Lastly, we implemented the GPU memory cleaning code that allowed us to train and use the model again.

Tasks for Next Week:

Next week, our focus will be on finalizing the model evaluation and completing the project report. We plan to conduct final tests on the trained model using our held-out test dataset. This will provide an unbiased evaluation of the model's generalization ability. We will compile all relevant performance metrics, including accuracy, precision, recall, F1-score, BLEU score, perplexity, token-level accuracy, and AUC-ROC, to thoroughly assess the model's performance. We will then finalize the project report, ensuring it includes a detailed description of the project, methodologies, results, a thorough discussion of our findings, the challenges we encountered, the solutions we implemented, and potential future work. The report will be well-structured, clearly written, and include all necessary visualizations and supporting information. In addition to finalizing the report, we will continue to refine the chatbot's functionality, focusing on handling edge cases, improving response quality, and enhancing the overall user experience.

Instructor's Feedback:

Instructor's Signature: _____
Date: November 29, 2024