Group 7 Project Proposal

Title: Developing an Intuitive GUI-Based Question Answering Bot using an LSTM Seq2Seq Model

Team Members:
- Manushi
- Ameya Santosh Gidh
- Narayana Sudheer Vishal Basutkar

Problem Statement:
In the current landscape of conversational AI, while there are numerous Question Answering Bots, there is a notable gap in providing a comprehensive solution that offers both a user-friendly graphical interface and robust conversational abilities. Our project aims to address this gap by developing an advanced Question Answering Bot that boasts an intuitive GUI, powered by a Sequence-to-Sequence (Seq2Seq) model built using TensorFlow's Long Short-Term Memory (LSTM).

Dataset Used:
Our project leverages the Stanford Question Answering Dataset (SQuAD), a widely recognized and popular dataset in the field of natural language processing and machine reading comprehension. SQuAD is designed to evaluate the ability of machine learning models to read a passage of text and provide accurate answers to questions about its content. This rich dataset comprises a diverse collection of paragraphs from various sources, including Wikipedia articles, news reports, books, and more. Each paragraph is associated with a set of questions and corresponding answers, making it a valuable resource for our research.

Approach to Solution:
Our approach to tackling this problem is comprehensive and well-structured:

1. Dataset Preprocessing: Data preparation is a critical initial step. We perform a series of text preprocessing tasks, including data cleaning to eliminate missing values, text normalization by handling contractions and special characters, and tokenization of both input and output sequences. Additionally, we create sequence pairs for our model.

2. Model Selection: Our model of choice is the Seq2Seq model, with its core NLP component being the Long Short-Term Memory (LSTM) architecture. The encoder-decoder framework plays a pivotal role in enabling the bot to comprehend and generate meaningful responses.

Related Work:
Our project draws inspiration from existing research in the field of Natural Language Processing and Seq2Seq models. Two influential papers that have guided our approach are:

1. Sequence to Sequence Learning with Neural Network (Sutskever et al., 2014): This seminal work discusses the effectiveness of Seq2Seq models, originally designed for machine translation, which laid the foundation for our choice of architecture.
- [Paper Link](https://arxiv.org/abs/1409.3215)

2. Incorporating Copying Mechanism in Sequence-to-Sequence Learning" by Jiatao Gu et al. (2016): This paper explores advanced sequence-to-sequence frameworks that could enhance the chatbot's capabilities. The concepts presented here have influenced our research.
- [Paper Link](https://arxiv.org/abs/1603.06393)

Assessment Methodology:
Our project's success will be assessed through a combination of performance evaluation measures, cross-validation strategies, and ablation studies:

- Performance Evaluation Measures:
1. Accuracy: The accuracy of the bot's responses will be a key performance indicator, ensuring it provides correct answers.
2. Computational Time: We will measure the time taken by the model to generate responses, which is crucial for real-time interaction.

- Cross-Validation Strategy:
- Data will be divided into train and test sets, enabling us to evaluate the model's performance while avoiding bias.

- Ablation Study:
- We will conduct an ablation study to fine-tune our model. This will involve varying hyperparameters such as learning rates and regularization parameters and experimenting with different configurations of LSTM cells, layers, and dropout rates to optimize our bot's performance.

Conclusion:
Our project represents a comprehensive effort to create a sophisticated Question Answering Bot. It excels in understanding and effective conversational capabilities while ensuring a delightful user experience through an intuitive graphical interface. By harmoniously merging these two critical aspects, we aim to elevate the utility and accessibility of question answering bots. Our project harnesses the power of state-of-the-art NLP models, meticulous data preprocessing, and established evaluation methodologies to deliver a valuable and user-friendly solution in the domain of conversational AI.