Capstone Project: NLP for Question Answer Chatbot

Objective:
The objective of this capstone project is to develop a question-answer chatbot using natural language processing (NLP) techniques. The project involves text data collection, preprocessing, model training using basic transformation models and different Large Language Model (LLM) architectures, evaluation, fine-tuning, and deployment of the best-performing model as a web application.

Dataset:
- The dataset for this project should comprise at least 20 documents.
- Text data collection will involve gathering documents relevant to the question-answer task.

Project Workflow:
1. Text Data Collection:
   - Gather at least 20 documents.
   - Organize the documents into appropriate categories.

2. Data Management:
   - Upload the text data to a cloud storage platform (e.g., SharePoint, Google Drive) for easy accessibility and collaboration.

3. Text Preprocessing:
   - Preprocess the text data to remove noise, such as punctuation, stopwords, and special characters.
   - Perform tokenization and lemmatization to standardize the text.

4. Model Building:
   - Build a basic transformer model for question-answering tasks.
   - Also, Apply different paid and open-source language model (LLM) architectures (e.g., BERT, GPT) for question-answering tasks.

5. Model Evaluation:
   - Evaluate the performance of each model using appropriate evaluation metrics.
   - Compare the performance of the basic transformer model and the LLM models.

6. Fine-Tuning LLM Models:
   - Fine-tune the selected LLM models on the specific question-answering task to optimize performance.

7. Model Selection:
   - Select the best-performing model based on evaluation metrics and fine-tuning results.
   - Justify the selection of the best model based on its performance and suitability for the question-answer chatbot.

8. Web Application Development:
   - Develop a web application to deploy the best-performing model as a question-answer chatbot.
   - Provide user-friendly interfaces for inputting questions and displaying corresponding answers.

9. Reporting:
   - Prepare a detailed text analysis and LLM modeling report documenting the project workflow, findings, and conclusions.
   - Create a PowerPoint presentation summarizing the key aspects of the project for presentation.

Deliverables:
- Organized dataset with at least 20 documents.
- Preprocessed text data.
- Jupyter Notebook or Python script containing code for model building, evaluation, and fine-tuning.
- Web application URL and files for deploying the question-answer chatbot.
- Detailed text analysis and LLM modeling report.
- PowerPoint presentation slides.

Evaluation Criteria:
- Completion of each project stage.
- Quality of text preprocessing, model building, and evaluation.
- Effectiveness of LLM model fine-tuning for question-answering tasks.
- Successful deployment of the web application with the question-answer chatbot functionality.
- Clarity and coherence of the text analysis and LLM modeling report.

Resources:
- Sample Dataset Source: [https://www.kaggle.com/]