Personalized Chatbot with Dynamic Responses

Project Description:

The project utilizes a text document containing business information about a ready-mix concrete company. The goal is to develop a chatbot that enables the company's employees to easily search and retrieve detailed business information from this document.

Group Members & Roles:

1- Aliaa Abdelkhalek Abdelwahab Hegazy (Leader)
2- Khaled Mohamed Aly Harb
3- Alaa Eid Sayed
4- Nesma Mohamed Abd Allah

Team Leader:

Aliaa Abdelkhalek Abdelwahab Hegazy

Objectives:

- To develop an intelligent chatbot capable of understanding and responding to employee queries related to the company's business information.
- To enable employees to quickly access detailed information stored within the company's text document.
- To improve information retrieval efficiency and reduce the time spent searching for business data.
- To demonstrate the integration of natural language processing techniques in business knowledge management.

Tools & Technologies:

- Python (for data manipulation and preprocessing).
- NLTK or SpaCy for tokenization, stop word removal, and lemmatization.
- Hugging Face Transformers (for transformer-based models like GPT and BERT).
- TensorFlow or PyTorch (for model training and fine-tuning), integrating attention mechanisms and implementing continuous learning techniques.
- Apache Airflow or Kubeflow for orchestrating the learning pipeline.
- MLflow for model tracking and versioning.
- Flask, FastAPI, or AWS API Gateway for deploying the chatbot as an API.

- GitHub Actions, Jenkins for CI/CD.

Milestones & Deadlines:

Milestone 1: Data Collection and Preprocessing (18-Oct)

Milestone 2: Chatbot Development and Training (01-Nov)

Milestone 3: Advanced Techniques and Integration (15-Nov)

Milestone 4: MLOps and Model Management (29-Nov)

Milestone 5: Final Report, Presentation, and Demonstration (05-Dec)

KPIs (Key Performance Indicators):

1. Data Preparation Quality

• Measures the completeness, cleanliness, and suitability of the dataset used for training (including preprocessing, tokenization, and splitting). **(Aliaa-Khaled)**

2. Model Performance & Accuracy

• Evaluates how well the trained model performs on key metrics (e.g., BLEU, ROUGE, Perplexity, FID, or human evaluation), showing how accurate or realistic its output is. **(Alaa-Nesma)**

3. Pipeline Integration & Automation Level

• Assesses how effectively all components (data preprocessing, training, inference, post-processing) are integrated into a single automated workflow or pipeline. **(Aliaa-Khaled)**

4. MLOps & Deployment Readiness

• Measures how well MLOps practices (tracking, versioning, CI/CD, deployment, monitoring) are implemented to ensure scalability and maintainability. **(Alaa-Nesma)**

5. Output Quality & Usability Score

• Evaluates the quality and usefulness of the generated content/output — text, chatbot responses, or generated images — from a human usability perspective. **(Aliaa)**

6. Documentation & Presentation Quality

• Assesses clarity and completeness of all reports, notebooks, code documentation, and the final presentation/demo. **(Nesma)**