

Round 4 – AI-ML Developer Intern

Objective: Build an offline chat-reply recommendation system using Transformers, trained on two-person conversation data.

Problem Statement:

You are given two separate datasets containing long two-person chat conversations. The task is to design a system that can predict the next possible reply from User A when User B sends a message — using User A's previous conversation history as context.

System Requirements:

- 1 Preprocess and tokenize long conversational data efficiently.
- 2 Fine-tune or train a Transformer-based model (e.g., BERT, GPT-2, or T5) offline.
- 3 Generate coherent, context-aware replies.
- 4 Evaluate responses using metrics like BLEU, ROUGE, or Perplexity.
- 5 Justify model choice, optimization, and deployment feasibility.

Proposed Methodology:

1. Load and preprocess data from userA_chats.csv and userB_chats.csv. 2. Tokenize using GPT-2 tokenizer for compatibility with language generation tasks. 3. Fine-tune a pre-trained GPT-2 model locally to generate context-aware responses. 4. Implement evaluation using BLEU score for generated vs actual replies. 5. Save the fine-tuned model for offline deployment using Joblib.

Tools and Libraries:

Python 3.10+, Transformers, Torch, Numpy, Pandas, Scikit-learn, Matplotlib, NLTK, Joblib.

Model Justification:

GPT-2 is selected for its ability to handle conversational context and generate coherent text. Its pretrained weights allow offline fine-tuning, making it ideal for this offline task.

Evaluation Metrics:

BLEU score was used to measure similarity between generated replies and actual responses. Additional qualitative checks ensured coherence and contextual relevance.

Results & Discussion:

The fine-tuned GPT-2 model achieved meaningful and contextually relevant responses for given user prompts. The system works efficiently offline with reasonable memory usage.

Conclusion:

The offline chat-reply recommendation system successfully generates context-aware replies using a fine-tuned GPT-2 model. It meets the evaluation and optimization criteria and demonstrates deployment feasibility.

Submitted by: Shipra Tiwari

File Name: ChatRec_Model.ipynb, Model.joblib, Report.pdf, ReadMe.txt