Internship Report: Medical Q&A Chatbot

Introduction

This project focuses on developing a Medical Q&A chatbot equipped with natural language processing and entity recognition capabilities. The chatbot leverages a combination of pre-trained AI models and user-friendly interfaces to provide accurate medical information. The project aims to enhance the accessibility of medical knowledge through interactive technology.

Background

The Medical Q&A chatbot integrates advanced AI models for natural language understanding and entity recognition. By employing state-of-the-art embeddings and vector search techniques, the system efficiently retrieves relevant medical information. The project addresses challenges in accurate medical data retrieval and interactive query handling.

Learning Objectives

1. Develop a chatbot capable of accurate medical Q&A using advanced AI techniques.
2. Gain experience in integrating vector databases and embedding models.
3. Understand the preprocessing of medical datasets for NLP tasks.
4. Enhance skills in building interactive Streamlit applications.

Activities and Tasks

1. Designed a chatbot interface using Streamlit for interactive medical queries.
2. Processed the MedQuAD dataset for structured medical QA pairs.
3. Integrated FAISS for efficient vector-based search and retrieval.
4. Implemented entity recognition using spaCy for enhanced query understanding.
5. Utilized pre-trained embeddings like PubMedBERT for contextual accuracy.

Challenges and Solutions

- **Challenge:** Processing large medical datasets efficiently.

  Solution: Optimized preprocessing pipelines and dataset handling.

- **Challenge:** Ensuring the chatbot provides accurate answers.

Solution: Leveraged pre-trained embeddings and vector-based retrieval.

- **Challenge:** Extracting medical entities from unstructured text.

  Solution: Employed spaCy for named entity recognition.

Outcomes and Impact

The Medical Q&A chatbot successfully demonstrates the potential of combining vector databases with advanced NLP models for interactive applications. The system efficiently processes and answers medical queries, showcasing practical applications in the healthcare domain.

Conclusion

This project highlights the importance of integrating advanced AI models with user-friendly interfaces to address real-world problems. By leveraging state-of-the-art technologies, the chatbot provides a robust platform for medical Q&A. This work serves as a foundation for further advancements in healthcare AI systems.