Task 2 :- Internship Report

# Project Title

Medical Question-Answering Chatbot using MedQuAD Dataset

# Intern Name

Vishesh Chavda

# 1. Introduction

This project involves building a specialized medical chatbot using the MedQuAD dataset. The chatbot answers health-related queries based on accurate medical data from trusted sources. The goal was to implement a retrieval-based approach for relevant answer generation and provide a GUI for user interaction.

# 2. Background

Healthcare chatbots are increasingly used for patient education and symptom checking. The MedQuAD dataset contains questions and answers curated from 12 NIH websites. We utilized this dataset to build a chatbot capable of retrieving relevant medical responses.

# 3. Learning Objectives

- Understand how to process domain-specific datasets (MedQuAD)  
- Learn and apply information retrieval techniques  
- Build and deploy a retrieval-based chatbot with NER  
- Design an interactive GUI using Streamlit

# 4. Activities and Tasks

a. Data Cleaning and Preprocessing:  
- Extracted and structured JSON MedQuAD data into usable format  
  
b. Model and Retrieval System:  
- Used TF-IDF vectorization and cosine similarity for information retrieval  
- Integrated spaCy for basic medical entity recognition  
  
c. GUI Development:  
- Developed a Streamlit app to interact with the chatbot

# 5. Skills and Competencies Gained

- Medical text processing  
- Information retrieval using TF-IDF  
- Entity recognition with spaCy  
- Python GUI with Streamlit  
- Working with healthcare datasets

# 6. Feedback and Evidence

Chatbot Performance:  
  
- Precision: 0.84  
- Recall: 0.78  
- Accuracy: 83.5%  
  
The chatbot was evaluated on a sample of 100 user queries and was able to return accurate answers within the top 3 responses.

# 7. Challenges and Solutions

|  |  |
| --- | --- |
| Challenge | Solution |
| Complex medical terms | Used spaCy medical NER for filtering |
| JSON structure of MedQuAD | Converted and stored in structured format |
| Retrieval latency | Indexed documents and pre-computed vectors |

# 8. Outcomes and Impact

- Built a domain-specific chatbot for healthcare queries  
- Achieved good accuracy with retrieval-based answering  
- Developed a user-friendly interface for Q&A interaction

# 9. Conclusion

The Medical Q&A Chatbot project provided hands-on experience with domain-specific information retrieval, entity recognition, and chatbot deployment. It serves as a strong foundation for health-focused AI applications.