

Fine-Tuning a Large Language Model for Medical Question Answering

Presented by: Atharva | Prompt Engineering



Made with GAMMA

Project Overview

Goal

Create specialized medical QA system using Flan-T5-base

Dataset

13 curated medical QA pairs covering diverse topics

Model

Google's Flan-T5-base with 250M parameters

Rationale

Small, high-quality data + instruction-tuned foundation model



Implementation Approach

Dataset Preparation

Formatted, structured medical QA pairs ready for fine-tuning

Training Configuration

- Learning rate: $1e-4$
- Batch size: 4
- Weight decay: 0.01
- Epochs: 10

Model Evaluation & Results

Qualitative Testing

Expert review of generated answers for accuracy

Example Q&A

Clinical questions with precise model responses

Limitations

Some inaccuracies & domain-specific gaps observed

Evaluation Style

Informal expert assessment without formal metrics



Inference Pipeline

Process

- Input question preprocessing
- Model inference
- Answer generation & output

Generation Parameters

- Temperature: controls randomness
- Beam search: improves answer quality

Interactive Interface

Screenshot showing user query input and generated reply

Limitations & Future Directions

Current Limitations

Small dataset, single hyperparameters, qualitative evaluation

Proposed Improvements

Expand dataset, add formal metrics, tune hyperparameters

Lessons Learned

Importance of data quality and careful evaluation

A graphic on a dark grey background. It features two orange arrows pointing upwards and to the right, one slightly above the other. Below the arrows, the text "AI MODEL" is written in large, white, bold, sans-serif capital letters. Below that, the word "Improvement" is written in a larger, white, bold, sans-serif font, with the 'I' and 'm' being significantly larger than the other letters. The entire graphic is set against a dark grey background.

AI MODEL
Improvement

Demo & Results Summary

Demo Overview

Shows model response generation to medical queries

Achievements

Functional specialized medical QA system using Flan-T5

Conclusion

Promising approach for medical LLM applications in education and research



Key Takeaways & Next Steps

1

Successful Fine-Tuning

Instruction-tuned Flan-T5 adapts well to medical QA

2

Data Quality is Crucial

High-quality curated pairs maximize performance

3

Future Work

Broader datasets, automated metrics, interface enhancements

4

Potential Impact

Assist medical education, reduce clinician workload