High Risk Project

Using LoRA Fine-Tuning of LLMs to Accurately Triage Patients

Cameron Morrongiello

Introduction

Topic and Importance

- Triage ensures patients receive the right level of care quickly.
- Overcrowded ERs and under-triaging put lives and resources at risk.
- NLP models could automate basic triage decisions.
- Goal: Fine-tune a small LLM to accurately suggest triage levels.
- Sub Goal: Create a synthetic dataset

Methods

Data Generation and Fine-Tuning

- Generated synthetic patient cases (symptoms, history, demographics, ie).
- Labeled cases into triage levels (ER, Urgent Care, Primary Care, Self-Care, ie).
- Fine-tuned SmolLM2-360M using LoRA
- Evaluated base vs. fine-tuned model on test set.

Results

Data Generation and Fine-Tuning

- Base model accuracy: 5%; Fine-tuned model: 62%.
- Base model answer rate: 33%; Fine-tuned: 100%.
- Fine-tuned model reliably produced and reasoned triage decisions.
- Demonstrates viability of lightweight LLMs for clinical support tasks.

Future Directions

Future Work and Expansion

- Improve complexity of synthetic dataset
- Fine-tune on real-world data (e.g., MIMIC-III clinical notes).
- Compare with medical-pretrained models (e.g., ClinicalBERT).
- Deploy as a nurse-assisting triage recommender system.

Demo