Jake Wendling, Samuel Amoabeng

Project 4 Proposal

Introduction

We aim to use few-shot learning with a pre-trained language model to diagnose a patient based on description of their symptoms.

Data Sources

The dataset we are using has "natural language descriptions of symptoms labeled with 22 corresponding diagnoses". We found the dataset on hugging face at https://huggingface.co/datasets/gretelai/symptom_to_diagnosis.

Methods and Techniques

- Pretrained language model
- Few shot learning

Deliverables

We aim to deliver a large language model that has been trained to give a diagnosis when given a few sentences that describe the symptoms of a patient. This will not be a catch-all model and will only have 22 diagnoses.