# **LLM Fine-tuning: Impression Generation**

Done by:

G. Preetham

AP21110010793

saipreetham\_grandhi@srmap.edu.in

## **Overview of Project**

This project will output medical impressions from radiology reports, presumably containing fields like "Report Name," "History," and "Observation." The goal is to fine-tune a pre-trained language model T5 for the medical impression generation task, evaluate it using ROUGE scores and perplexity, and visualise the text-based patterns in the dataset as a similarity graph.

**Approach and Methodology**

**Data Preprocessing:**

**Data selection**: We first selected the columns appropriate for the dataset. These columns were "Report Name," "History," "Observation," and "Impression." These columns make up the input-output pairs to train.

**Text Cleaning**: As a prerequisite to feeding the text into the analysis and feeding it into the model, the following text preprocessing steps were executed:

stops word deletion with NLTK's list of English stopwords

text normalization by stemming and lemmatization

combining the text inputs built by combining the "Report Name," "History," and "Observation" fields.

**Train Test Split**: The dataset was split into a training set with 300 samples and an evaluation set with 30 samples. We used these splits for model training and testing.

**TF-IDF and Cosine Similarity**

We made use of Term Frequency-Inverse Document Frequency to vectorize the lemmatized text so we can explore the data and identify patterns.

We calculated the cosine similarities between words so we know which word pairs are most similar across the dataset.

A graph was created to illustrate the top 100 word pairs in terms of similarity, where edges represent word association.

**Model Fine-tuning**: The model was a T5 model pre-trained on general text data, which will then be leveraged to generate medical impressions

**Input to the model:** Concatenation of "Report Name," "History," and "Observation" fields

**Output from the model:** "Impression" fields.

For every evaluation sample, the model created an impression and compared this with the actual impression.

**Evaluation Metrics:**

**ROUGE Scores:** In order to measure how closely the generated medical impressions resembled the actual impressions, we used ROUGE-1, ROUGE-2, and ROUGE-L metrics that define overlap between the generated text and the target text at different n-gram levels.

**Perplexity Score**: Perplexity was calculated for how well the model predicts the next token in a sequence. A low perplexity is a sign of a good model.

**Visualization:** A network graph was designed to visualize relationships that existed between the most similar words in the dataset by cosine similarity scores. This type of visualization ensured that the co-occurrence of words and thematic patterns in radiology reports have been taken into account.

Assumptions Made

**Input Fields Relevance:** Assuming there is enough detail in "Report Name," "History," and "Observation" to synthesize medical impressions, concatenating these fields would be a meaningful representation of the radiology report.

**Pre-trained Language Model (T5):** We assume that the T5 model pre-trained on general datasets can be further fine-tuned on this particular domain, namely medical impressions, with no special domain-specific pre-training.

**Data Cleaning:** During text pre-processing, some words were removed based on the assumption that stopwords did not hold useful information in the context of the medical report.

**Textual Similarity**: Any cosine similarity between TF-IDF vectors was considered to carry useful insights into the relationships between words in medical text, which potentially portrayed conceptual similarities between terms.

**Conclusion**

In the generation and evaluation of medical impressions from radiology reports, preprocessing, model fine-tuning, and evaluation techniques are utilized. More importantly, key metrics used for assessing performance levels include ROUGE and perplexity measures that can be used as a baseline to further refine and develop more impressive models.