MAI 475– Large Language Model
IV MSAIM
26-06-2024

Regular lab Question

Lab Exercise 3:

Implement a Comparative Analysis of Foundation and Domain-Specific Models

Comparative Study on any three models: (1 foundation and 2 Domain-Specific)

- Understand the importance of Domain-Specific models in various domains (**Finance, Education, Healthcare, Legal, Sports, E-commerce, News & Media and more)**
- Compare and contrast different transformer-based foundation models such as **GPT**, **BERT, T5**, **BART**, **FLAN-T5** or other with domain-specific models **Model Selection**
- **Domain-Specific Model** Examples:
  Finance: FinBERT, BloombergGPT, **Healthcare**: BioGPT, ClinicalBERT
  Legal: Legal-BERT, CaseLaw-BERT, **Education**: EduBERT, SciBERT
  Sports/Media: MediaSum-based T5, **E-commerce**: AmazonProductBERT

Dataset and Preprocessing (Optional):

- Choose an appropriate dataset or create synthetic dataset.
- Explain your data preprocessing steps such as tokenization, prompt formatting, and truncation or padding.

Evaluation and Results:

Expected Output and Performance Evaluation

Output: (Choose Suitable Visualization)

- Generate **minimum 3 comparative outputs** using both models.

- Provide **screenshots or logs** of outputs.

Evaluation Metrics: (Select the suitable PEM for the corresponding task)

- BLEU / ROUGE for text quality

- Human evaluation: Coherence, relevance, factual accuracy

NB: No marks will be credited for the pipeline implementation.

Program Evaluation Rubrics

Model Selection and Implementation	6 Marks
Timely Submission	2 Marks
Viva	2 Marks

General Instructions

- The file you have to save with your name, last 3 digits of register number and program number "Anto_501_Lab1".
- The implemented code you have to download and upload in the Google Class room in the given scheduled time.