University of Peradeniya

Generating Patient Clinical Letters Using LMs

M. Nanayakkara[1], M. De Alwis[1]

Supervised by: A. Bandaranayake[1], D. Hearth[1], R. Ragel[1], L. Ariyasinghe[2]

[1]Department of Computer Engineering, Faculty of Engineering, University of Peradeniya. [2]Axiata Digital Labs (Private) Limited

Abstract - This study explores using advanced Language Models (LMs) to automate patient clinical letter generation, aiming for efficiency and clarity. It evaluates SLMs for accuracy and linguistic precision, seeking the best model for generating these letters while addressing data privacy concerns.

Background:

Patient clinical letters are crucial in healthcare, offering a comprehensive summary of findings, diagnoses, and treatment plans for patients and healthcare providers. These letters facilitate effective communication within the healthcare team and empower patients to make informed decisions about their health.

Objective:

Language models are divided into large (LLMs) and small (SLMs) categories. SLMs excel in efficiency, precision, and ease for generating patient letters, offering cost-effective solutions for resource-limited healthcare. They require less data for targeted training, ensuring confidentiality of patient information while enhancing accuracy and security.

Research Focus:

This research aims to evaluate the performance of SLMs and LLMs in generating patient clinical letters. By comparing their efficiency, accuracy, and security features, we seek to identify the most effective model for enhancing healthcare documentation and communication.

Materials & Methodology

- *Objective:* Investigate the feasibility of using Small Language Models (SLMs) for generating patient clinical letters, comparing their performance with Large Language Models (LLMs) and ChatGPT.
- *Dataset and Prompts:* Utilized a brain tumor-related dataset to generate prompts for clinical letters.
- *SLMs Tested:* Tested SLMs including Phi3, Tiny Llama, Zephyr, and Llama3 against LLMs and Chat GPT in generating discharge, referral, and consultation letters related to brain tumors.
- *Dataset Creation:* Generated 30 letters for each SLM, covering various clinical scenarios and complexities in brain tumor cases.
- *Evaluation Criteria:* Compared letters for readability, inference time, and accuracy against those generated by LLMs and ChatGPT.
- *Readability Assessment:* Used WebFX tool for evaluating readability.
- *Inference Time Measurement:* Employed Ollama and Flask to record the time taken for letter generation.
- *Further Training:* SLMs further trained using a ChatGPT-generated dataset to improve accuracy and readability.
- *Comparison with ChatGPT:* Compared accuracy and readability of generated clinical letters with Chat GPT to determine SLMs' performance compared to LLMs.

Results

Tiny llama has the minimum amount of parameters which is 1/8 times from other 3. According to the results tiny llama has the highest readability and also the minimum inference time. Since phi 3 and tiny llama has the least amount of word generation and good readability score we have decided to use those 2 slms for further training.

	Readability score (%)			
Output	llama3	phi3	tiny llama	Zephyr
Average Readability Score	34.87	30.04	41.67333333	38.89666667

	Inference Time (seconds)			
Output	llama3	phi3	tiny llama	Zephyr
Average Inference Time	144.5333667	55.6118	15.44	94.44540667

	Content Length (words)			
Output	llama3	phi3	tiny llama	Zephyr
Average Content Length	386.5	470.9666667	363.6333333	319.4

	Word Generation Time (seconds/words)			
Output	llama3	phi3	tiny llama	Zephyr
Average Time	0.3742981519	0.1177338248	0.04173401912	0.2946395331

Contact details

Name : Muthuni De Alwis

Tel. No.: +94771016720

Email : e18058@eng.pdn.ac.lk

Multidisciplinary AI Research Centre (MARC)
University Research Council
University of Peradeniya
Peradeniya, 20400, Sri Lanka