FRI Natural language processing course 2025
Project report

University *of Ljubljana*
Faculty *of Computer and*
Information Science

Automatic generation of Slovenian traffic news for RTV Slovenija

Vid Kobal, Matija Tomažič, and Žan Luka Nusdorfer

Abstract

This project aims to fine-tune an existing large language model (LLM) for the task of generating concise and relevant traffic news reports. The goal is to automate the production of traffic updates for radio broadcast, based on structured data from the promet.si portal, a task currently being done manually by students at RTV Slovenija. We use various prompt engineering techniques and evaluate the effectiveness of outputs with (semi) automatic and human evaluation metrics. We take the following into account; accurate road naming, filtering of relevant events, correct filtering, text lengths and words and editorial guidelines. The report also discusses the preparation of datasets, challenges of generation in a low-resource language, and the design of an interface for interactive testing of model outputs.

Keywords
prompt engineering, text generation, traffic news, fine-tuning, data-to-text generation, RTV Slovenija

Advisors: Slavko Žitnik

Introduction

In an effort to modernize and automate the generation of traffic news for RTV Slovenija, this project explores the application of natural language processing (NLP) techniques for the automatic generation of Slovenian traffic reports. The main goal is to use a large language model (LLM), fine-tune it for this specific use case and to use prompt engineering for producing short and relevant traffic updates based on structured data from the promet.si portal. These reports are intended for radio presenters who regularly read them to inform the public.

The task is currently being done manually. While functional, this approach is more labor-intensive, prone to variability, and limited in terms of scalability. By automating the process, we hope to improve efficiency, consistency, and content quality, while following the editorial standards for language style, relevance, and format.

One of the things we focus on is the correct identification of relevant traffic events, accurate road naming, and appropriate text length and filtering. We achieve this with prompt engineering and parameter-efficient fine-tuning (e.g., LoRA). We evaluate generated text using automatic (precision, recall, F1) as well as human evaluation metrics. The project also touches on the design of interactive tools for testing model outputs, and explores best practices in dataset preparation and model evaluation.

In summary, this project aims to deliver a reliable and efficient system for automated traffic news generation by leveraging Large Language Models (LLMs), their fine-tuning, and prompt engineering techniques.

Related work

Generating informative traffic new from structured data falls within the broader domain of data-to-text (D2T) generation, specifically under table-to-text generation. For our project, structured traffic data is supplied by the promet.si portal in CSV format. The data includes fields such as road names, event types, timestamps and locations. The objective is to transform these data entries into concise and reliable reports that are ready for radio broadcasters to read.

Recent research has demonstrated that LLMs, when combined with data cleaning and model fine-tuning, are highly effective for table-to-text generation. For example, the HeLM (Highlighted Evidence augmented Language Model) framework [1] proposes a two-step approach, where a highlighter selects more important rows. The highlighter is then followed by a summarizer that produces human-readable text.

This method achieved state-of-the-art BLUE (bilingual evaluation understudy) and ROGUE (Recall-Oriented Understudy for Gisting Evaluation) scores, confirming that pre-selecting important data improves both model performance and text understandability.

Using LLMs in low resource environments requires parameter-efficient fine-tuning. Methods like LoRA (Low Rank Adaptation) [2] and QLoRA (Quantized Low Rank Adaptation) [3] allow fine-tuning models on modest hardware while maintaining high accuracy. Researchers have also found that smaller, fine-tuned models ensure reliability and resource efficiency [4].

Another essential part of our project is exploring and testing different prompt engineering techniques. This involves carefully crafting prompts to elicit specific, high-quality outputs from LLMs. Many studies have been done on prompt engineering techniques, in which researches test different prompt engineering techniques (zero-shot and few-shot prompting, chain-of-thought reasoning, and self-consistency decoding) [5]. In traffic news generation, these strategies help guide the model, while also helping it avoid unnecessary or irrelevant information.

Practical guidance for prompt engineering is provided by Prompting Guide AI (https://www.promptingguide.ai/), which offers extensive explanations and specific advice on different prompt engineering approaches. In addition, Ollama has demonstrated effective prompt workflow with models like Code Llama, especially for tasks involving structure data transformation. They offer different examples on how to prompt Code Llama to optimize relevance and clarity (https://ollama.com/blog/how-to-prompt-code-llama).

In conclusion, our project builds upon this body of research in table-to-text generation, fine-tuning methods, and prompt engineering. By integrating structured data processing, efficient LLM adaptation, and strategic prompt design, we aim to deliver automated and reliable compliant traffic news that meets the standards of RTV Slovenija for relevance, clarity, and factual accuracy.

References

[1] Junyi Bian, Xiaolei Qin, Wuhe Zou, Mengzuo Huang, Congyi Luo, Ke Zhang, and Weidong Zhang. Helm: Highlighted evidence augmented language model for enhanced table-to-text generation, 2024.

[2] Yuchen Zeng and Kangwook Lee. The expressive power of low-rank adaptation, 2024.

[3] Yelysei Bondarenko, Riccardo Del Chiaro, and Markus Nagel. Low-rank quantization-aware training for llms, 2024.

[4] Joy Mahapatra and Utpal Garain. Impact of model size on fine-tuned llm performance in data-to-text generation: A state-of-the-art investigation, 2024.

[5] Shubham Vatsal and Harsh Dubey. A survey of prompt engineering methods in large language models for different nlp tasks, 2024.