



Automatic generation of Slovenian traffic news for RTV Slovenija

Anže Javornik, Timotej Rozina, Anže Šavli

Abstract

Keywords

traffic news, automatic generation, RTV Slovenija, news generation

Advisors: Slavko Žitnik

Introduction

In this assignment, we will develop an automated solution for generating Slovenian traffic news reports tailored specifically for RTV Slovenija's radio broadcasts. The goal is to replace the existing manual process with an efficient, accurate, and automated system leveraging advanced language modeling techniques (LLMs).

We will start by reviewing existing research to choose the best LLM for the task and study the provided traffic data. Initially, we will use prompt engineering methods to optimize our results and set up clear evaluation criteria to check if the news reports correctly highlight important events, use accurate road names, follow formatting rules, and have the right length and clarity. Afterwards, we will fine-tune the chosen LLM for better performance and finally assess how well the automatic system performs compared to manual reports, using both numbers (like precision, recall, and F1-score) and human feedback.

1. Methods

1.1 Existing solutions

In this section we will look at some existing solutions that tackle similar challenges include several advanced and practical implementations currently in operation.

- GEWI's TIC Software is a prominent solution used by both commercial and government organizations for managing traffic data. This system collects, generates, and distributes various traffic-related information, such as incidents, weather conditions, and parking availability. It facilitates operators to manually input data or automate data integration from live traffic sources [1].

- Another innovative solution, TrafficGPT, employs multiple AI agents to generate comprehensive traffic information. It includes a text-to-demand agent for user interaction, a traffic prediction agent that analyzes temporal traffic data, and a visualization agent presenting traffic insights clearly [2].
- Waze offers another practical example with its voice-activated incident reporting feature. Users can verbally report traffic incidents, and the AI system within the app interprets and processes these reports, ensuring real-time and accurate updates [3].

1.2 LLM options

In this section, we will describe selected few LLM systems that we feel like would offer optimal performance for our task and goals.

- LLaMA - family of models, developed by Meta, stands out as a prominent open-source choice. The latest iteration, LLaMA 3, boasts improved efficiency, versatility, and a significantly large context window. This expanded context window could be particularly beneficial for processing larger datasets or more intricate reporting instructions, allowing the model to consider more information at once and potentially generate more comprehensive and contextually relevant reports. LLaMA 3 can be utilized locally.
- Mistral AI - offers set of open-source LLMs, including Mistral 7B and Mixtral. Mistral provides a fine-tuning API through its La Plateforme, simplifying the process of customizing these models. This dedicated API makes

it easier for developers to programmatically fine-tune Mistral models.

- Gemma - a family of lightweight open-source LLMs from Google, presents another compelling option. These models are available in different sizes, ranging from 2 billion to 27 billion parameters, offering flexibility to choose a model that aligns with available computational resources and desired performance levels. Gemma has shown strong performance compared to other open-source models and supports integration with various tools and frameworks.
- Anthropic's Claude - models are positioned as strong models with a particular emphasis on being helpful, honest, harmless, and safe for enterprise applications. Its ability to handle complex queries, potential for data analysis tasks (categorization, classification, and summarization of large datasets) and multi-step reasoning could be beneficial for generating detailed reports.

1.3 Initial ideas

To effectively develop an automated system for generating Slovenian traffic news for RTV Slovenija, we will initially experiment with prompt engineering techniques and test them on listed LLM options. This will involve crafting and testing various prompts to determine their effectiveness in producing clear, accurate, and contextually relevant traffic reports.

After establishing baseline results with prompt engineering, we will define clear evaluation criteria, such as accuracy in identifying important events, correct usage of road names, appropriate report length, and adherence to RTV Slovenija's broadcasting guidelines.

Finally, based on the insights gathered from these initial tests, we will plan the subsequent fine-tuning process for the selected LLM to enhance its performance further.

1.4 Dataset

Our dataset contains Slovenian traffic information for the years 2022, 2023, and 2024. Each year's data is organized in separate sheets. The dataset includes entries with details such as timestamps, operator names, and various categories of traffic information like weather conditions, roadwork, restrictions for heavy vehicles, general notices, and other road obstructions.

Results
Discussion
Acknowledgments
References

[1] GEWI. Gewi traffic information center (tic) software. <https://en.wikipedia.org/wiki/GEWI>. Accessed: 2024-03-21.

[2] Yuantong Sun, Xiangjie Yang, Bowen Li, and Chen Zhang. Trafficgpt: Multi-scale traffic generation with ai agents. *arXiv preprint arXiv:2405.05985*, 2024. Accessed: 2024-03-21.

[3] Andrew J. Hawkins. Waze now lets you report incidents by using your voice. <https://www.theverge.com/2024/10/31/24284037/waze-report-incidents-voice>, 2024. Accessed: 2024-03-21.