



Natural language processing course Latex Template

John Doe, Jane Doe, and Mike Smith

Abstract

Keywords

Keyword1, Keyword2, Keyword3 ...

Advisors: Slavko Žitnik

Introduction

Our initial goal is to leverage an open-source large language model (LLM) capable of generating text in Slovenian to automate the process of traffic report generation for RTV SLO. By doing so, we aim to streamline and enhance the efficiency of producing structured reports while ensuring consistency in formatting and adherence to predefined guidelines.

To achieve this, we plan to train the LLM to learn the specific formatting and stylistic requirements from existing data. This will allow the model to generate reports that align with the required standards without the need for extensive manual editing. Since using a large-scale model can be resource-intensive, we also intend to fine-tune a smaller, more efficient LLM tailored to our specific use case. This smaller model would retain the necessary capabilities while reducing computational costs, making it more practical for real-world deployment.

If time permits, we aim to extend the project by integrating automatic translation features for languages commonly spoken in Slovenia, such as English, German, and Italian. This would make the generated reports more accessible to a wider audience. As a potential final step, we may also explore the implementation of text-to-speech functionality, particularly for English, to further enhance accessibility and usability.

Through this approach, we seek to develop a scalable and efficient solution for automated report generation that meets the needs of Slovenian-speaking users while incorporating multilingual and accessibility features as an added value.

Related works

This work focuses on the automatic generation of traffic reports from traffic data. This intersects with several areas of research, including data-to-text generation, natural language

processing for traffic information, and the application of large language models (LLMs).

Data-to-Text Generation: The task of automatically generating textual descriptions from structured data has been a long-standing research area [1]. Our work contributes to this field by focusing on the specific domain of traffic data, aiming to produce informative and concise reports. Recent advancements in LLMs have significantly impacted data-to-text generation, offering new possibilities for creating more natural and contextually relevant outputs [2].

Traffic Information Processing with NLP: Several works have explored the use of NLP techniques for processing and extracting information from traffic-related sources. For instance, [3] investigates empowering real-time traffic reporting systems using NLP-processed social media data. Our work complements this by focusing on generating reports directly from structured traffic data. The automation of traffic incident management using loop data has also been explored recently [4].

Large Language Models for Text Generation and Understanding: The rise of large language models has opened new avenues for various natural language processing tasks, including text generation [5, 6, 7, 8]. Techniques such as prompt engineering [9] and fine-tuning [10, 10] are crucial for adapting these models to specific domains and tasks. Furthermore, the ability of LLMs to understand and utilize document layout for enhanced performance has been investigated [11]. While our work primarily focuses on generating text from structured data, these advancements in LLM capabilities are highly relevant.

Evolution of Collective Behavior and Linguistic Systems: Although seemingly distant, the foundational work on the evolution of collective behavior using linguistic fuzzy rule-based systems [12] and the promotion of parallel movement

through balanced antagonistic pressures [13] provide insights into the development of complex communicative behaviors in artificial systems, which can indirectly inform the design of more sophisticated traffic reporting systems in the future.

Article **A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT** explores possibilities of formatting prompts for existing LLMs, guiding them to ensure certain form and rules of the generated output. This could be used for the first part of the task, where our task would lead us to initially use prompt engineering for traffic report generation.

Methods

References

- [1] Microsoft Research. Data2text: Automated text generation from structured data, 2025. Accessed: 2025-03-19.
- [2] LlamaIndex. Building blocks of llm report generation: Beyond basic rag, 2025. Accessed: 2025-03-19.
- [3] Xiangpeng Wan, Michael Lucic, Hakim Ghazzai, and Yehia Massoud. Empowering real-time traffic reporting systems with nlp-processed social media data. *IEEE Open Journal of Intelligent Transportation Systems*, 1:159–175, 01 2020.
- [4] Matteo Cercola, Nicola Gatti, Pedro Huertas Leyva, Benedetto Carambia, and Simone Formentin. Automating the loop in traffic incident management on highway, 2025.
- [5] Domen Vreš, Martin Božič, Aljaž Potočnik, Tomaž Martinčič, and Marko Robnik-Šikonja. Generative model for less-resourced language with 1 billion parameters, 2024.
- [6] Wenhao Zhu, Hongyi Liu, Qingxiu Dong, Jingjing Xu, Shujian Huang, Lingpeng Kong, Jiajun Chen, and Lei Li. Multilingual machine translation with large language models: Empirical results and analysis, 2024.
- [7] Elijah Pelofske, Vincent Urias, and Lorie M. Liebrock. Automated multi-language to english machine translation using generative pre-trained transformers, 2024.
- [8] Xirui Peng, Qiming Xu, Zheng Feng, Haopeng Zhao, Lianghao Tan, Yan Zhou, Zecheng Zhang, Chenwei Gong, and Yingqiao Zheng. Automatic news generation and fact-checking system based on language processing, 2024.
- [9] Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C. Schmidt. A prompt pattern catalog to enhance prompt engineering with chatgpt, 2023.
- [10] Mathav Raj J, Kushala VM, Harikrishna Warriar, and Yogesh Gupta. Fine tuning llm for enterprise: Practical guidelines and recommendations, 2024.
- [11] Marcel Lamott, Yves-Noel Weweler, Adrian Ulges, Faisal Shafait, Dirk Krechel, and Darko Obradovic. Lapdoc: Layout-aware prompting for documents. In Elisa H. Barney Smith, Marcus Liwicki, and Liangrui Peng, editors, *Document Analysis and Recognition - ICDAR 2024*, pages 142–159, Cham, 2024. Springer Nature Switzerland.
- [12] Jure Demšar and Iztok Lebar Bajec. Evolution of Collective Behaviour in an Artificial World Using Linguistic Fuzzy Rule-Based Systems. *PLoS ONE*, 12(1):1–20, 2017.
- [13] Jure Demšar, Erik Štrumbelj, and Iztok Lebar Bajec. A Balanced Mixture of Antagonistic Pressures Promotes the Evolution of Parallel Movement. *Scientific Reports*, 6, 2016.