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

# Automatic generation of Slovenian traffic news for RTV Slovenija

Anže Hočevar in Jan Anžur

**Abstract**

This project investigates the automatic generation of Slovenian traffic news using large language models (LLMs), aiming to emulate the structure and content style of RTV Slovenija bulletins. We propose and compare three data preparation pipelines—basic, DP1, and DP2—which align structured traffic events from `promet.si` with reference summaries from RTF reports. The GaMS-27B-Instruct model is fine-tuned using these pipelines via instruction-tuning and LoRA adaptation. Evaluation includes lexical overlap, edit distance, and semantic similarity metrics, alongside API-based LLM scoring. Results indicate that while DP1 yields the most structurally faithful outputs, basic and DP2 variants are often preferred by LLM evaluators due to their greater flexibility. The findings suggest a trade-off between rigid formatting and content adaptability, with implications for both training strategy and prompt design in multilingual, domain-specific text generation.

**Keywords**

Large language model, fine-tuning, data processing

*Advisors: Slavko Žitnik*

## Introduction

Automated news generation (often called automated journalism or robot journalism) refers to software producing news from data with minimal human input [1]. Early NLG systems in newsrooms were largely template-based, using predefined phrases filled with data (e.g., finance or sports reports) [1]. Recent advances in deep learning and especially the Transformer architecture have greatly improved the ability to generate coherent, contextually accurate narratives from structured data [2]. For example, Leppänen et al. (2017) developed a data-driven NLG system that produced thousands of localized election news articles in multiple languages [1]. These systems highlight the potential of automated content creation, but also underscore requirements like transparency, accuracy, and adaptability in journalistic contexts [1]. Ensuring factual correctness and neutrality is paramount – research notes that maintaining objectivity and avoiding bias remain significant challenges for AI-generated news [3]. Modern LLMs, with their ability to generate fluent text, are now being explored as core engines for such NLG tasks.

## Literature Review

One of the key considerations in automating Slovenian traffic news is the availability and adaptability of LLMs for multilingual text generation. Studies have shown that most state-of-the-art models, including OpenAI's GPT series [4], BLOOM [5], and mT5 [6], demonstrate strong multilingual capabilities. However, Slovenian, being a low-resource language, remains under-represented in large-scale training corpora [7]. Locally fine-tuned models such as SloT5 [8] have emerged to address this gap, showing promise in domain-specific Slovenian text generation.

A crucial challenge in this domain is balancing between **fine-tuning** and **prompt engineering**. Fine-tuning LLMs on domain-specific text can improve accuracy but requires computational resources and well-annotated data [9]. Prompt engineering, on the other hand, provides a lighter-weight alternative by designing effective input prompts to guide the model's response [10]. For structured and time-sensitive content like traffic news, a hybrid approach may be necessary, combining a base model with well-optimized prompting techniques.

Additionally, context-aware traffic reporting can benefit from external data sources, such as live weather updates, public holiday schedules, and road congestion analytics. Studies

have indicated that integrating real-time sensor data and probabilistic event modelling improves the predictive accuracy of automated reports [11].

Based on these insights, the following sections analyse the provided structured dataset of Slovenian traffic news, examining patterns in report generation and urgency to inform an AI-based automation approach.

## Methods

### Data

We are working with traffic news data. More specifically, we have a set of traffic news reports that serve as a part of our training data. The other part is hidden in the periodically scraped website data that in some way or another corresponds to parts of the traffic news reports. The input to our NLP model will be unseen scraped website data, whereas the output will be a report that follows the intended format.

#### Models

We used an open-source large language model (LLM) named GaMS-27B-Instruct[1] to generate new data. Since our data is basically completely in Slovenian, we needed an LLM that was trained specifically in this language. Technically speaking, it is a fine-tuned version of the Google Gemma2 LLM[2], so it performs well both on English and Slovenian.

#### Data preprocessing variants

We chose two implement two different preprocessing pipelines:

- - Data preprocessing 1 (DP1), which creates input-output pairs of reports by taking an output (given a date) and pairing it with a flattened-input using multiple temporally-close inputs
- - Data preprocessing 2 (DP2), which splits the reports into paragraphs and tries to make input-output pairs using various NLP techniques for matching.

Both of these aim to take the given data and convert it into a specific format: a list of input-output pairs that we would like the LLM to "learn from".

#### Preprocessing Pipeline (DP1)

To align structured traffic logs from `promet.si` with RTV Slovenija's RTF reports, we designed a multi-step preprocessing pipeline:

**1. Temporal Filtering.** We extract all structured events within a 1–8 hour window prior to the RTF timestamp, ensuring coverage of both new and persistent incidents.

**2. Cleaning and Normalization.** All fields are cleaned using HTML stripping, whitespace collapsing, and timestamp normalization to `datetime`.

**3. Sentence Extraction and Deduplication.** Each row is split into sentences and deduplicated using:

- **Exact match:** Removes literal duplicates.
- **Semantic match:** Uses Sentence-BERT with cosine similarity $> 0.7$.

**4. Content Selection.** We retain key sentences via:

- Longest informative sentence,
- TF-IDF scoring,
- Named entity filtering (e.g., roads, locations).

**5. Input Formatting.** Selected sentences are flattened into either:

**(a)** Plain concatenation (minimal preprocessing), or

**(b)** RTV-style header with "`Prometne informacije DD.MM.YYYY HH.MM`" and "`Podatki o prometu.`" prefix.

**6. RTF Matching.** The closest-in-time RTF is parsed via regex to extract the reference summary.

This pipeline produces coherent model inputs structurally aligned with real RTV outputs and supports both evaluation and fine-tuning.

#### Preprocessing pipeline (DP2)

This pipeline is somewhat simpler in nature. It takes a given output and finds its best possible corresponding input. The stages include:

- Split given report (output) into individual events.
- Find and split temporally-close input data (from excel).
- Compare each event from the given report with all events from the input.
- Create input-output pairs using matches that meet a confidence threshold.

**1. Splitting into events.** After visually observing the dataset, we observed that the outputs (individual RTF files) are split into multiple paragraphs and that each paragraph corresponds to one event or a group of related events. In other words, the contents of different paragraphs pertain to different news stories. Additionally, each line in the input data (excel line) consist of multiple columns that can be (carefully) merged. After being merged, we notice that this data can also be split into events by the html tags. After this stage, we split every data point in the input and output into a series of events.

**2. Finding temporally-close input data.** This stage is fairly similar to its analogue from DP1. We simply take all the input data from a time-window around the time that the output data was obtained and split it into a flat list of events (outputs).

[1] https://huggingface.co/cjvt/GaMS-27B-Instruct
[2] https://huggingface.co/google/gemma-2-27b

**3. Matching inputs to a given output.** After the previous stage, we have a single output event and a list of candidate input events. We implemented a multi-stage paragraph-comparison procedure. This procedure takes into account the number of matching words, proper nouns (PN) and named entities (NE). The latter two have a significant overlap since they both roughly refer to a specific person, location, object etc., but are not the same in their implementation. We searched for pairs of matching words using Levenshtein distance on the lemmatized paragraphs (maximum distance of 1 for a match). Any given input-output pair (of events) had to pass a series of trials, where it would need to pass an adjusted minimum threshold for the number matching words, PNs and NEs. Finally, the pair needed to also achieve a high enough score using vector embeddings comparison.

## Baseline results

As a starting point, we attempted to approach this task by simply using just common prompt engineering methods. To put it more bluntly, we fed the LLM with a prompt that concisely and clearly describes the task at hand and provided a sequence of shots (examples) that it should be able to follow. An example of this approach in action is shown in figure 1. It is apparent that the output is very similar to the input. Although, this is not necessarily bad, as long as the rules for data generation are not violated.

## Fine-Tuning: Training Data Generation and Model Adaptation

To fine-tune our large language models (LLMs) on domain-specific Slovenian traffic news reports, we prepared a high-quality set of input–output training examples derived from publicly available sources. This pipeline is computationally intensive due to large-scale RTF parsing, semantic deduplication, and format conversion.

**1. Input–Output Pair Generation.** These pairs are stored in JSONL format for training with one object per line: { "input": "...", "output": "..." }. We convert structured excel logs and corresponding RTF summaries into training examples in the format:

DP1:
VHOD: <flattened, cleaned input paragraph >
IZHOD: <RTF-sourced summary matching the input's timestamp context (ground truth) >

DP2:
VHOD: <the input part of a match >
IZHOD: <the output part of the match >

**2. Prompt Construction.** We then formatted these pairs into an instruction-tuning template. It includes:

A task description (e.g. "You are a traffic report writer..."),
Three few-shot examples,
The current input as the final prompt, expecting model completion.

The style would be similar to what the LLM would recieve as input for inference after being fine-tuned.

### 3. Training Setup.

Models such as GaMS-2B, GaMS-9B-Instruct and GaMS-27B are fine-tuned using PEFT (LoRA) on each of the data preparation techniques, leveraging Hugging Face pipelines and mixed-precision execution. Inference outputs are saved in outputs.jsonl for evaluation.

**LoRA Configuration.** The model was fine-tuned using Low-Rank Adaptation (LoRA) with the following parameters:

- **`task_type` = `CAUSAL_LM`** — Targets causal language modeling tasks.
- **`r` = 64** — Rank of the low-rank adapter matrices.
- **`lora_alpha` = 64** — Scaling factor applied to the adapter outputs.
- **`lora_dropout` = 0.1** — Dropout probability applied to adapter layers.
- **`bias` = `"none"`** — Bias terms are not adapted.

## Evaluation

To assess the quality of model-generated outputs, we conducted inference across three distinct data preparation pipelines: *Basic*, *DP1*, and *DP2*. Each setting was evaluated over 20 randomly selected samples.

The **Basic prompt** consists of raw traffic event descriptions concatenated with minimal preprocessing and embedded HTML tags. An example of the generated output under this configuration is shown in Figure 1. As a simple enhancement, we then applied basic text cleaning—stripping HTML tags and collapsing whitespace—which resulted in improved coherence and stylistic alignment with human-written summaries. This cleaned variant is illustrated in Figure 2.

To further increase relevance and faithfulness to human reports, we developed two structured pipelines:

- **DP1** leverages temporal filtering of structured data based on timestamp alignment with a single target RTF report. The result is a longer, context-rich prompt paired with a verified human-written summary, enabling direct output–reference comparison (Figure 3).
- **DP2**, in contrast, doesn't necessarily take the whole output RTF into account. It matches paragraphs from it to paragraphs from a flattened list of inputs. While this makes precise reference-based evaluation more difficult, the generated output (Figure 4) remains interpretable and suitable for qualitative inspection.

**Table 1.** Summary of evaluation metrics on generated vs. ground-truth traffic reports

| Metric | Median | Std. Dev. |
|---|---|---|
| F1 Token Overlap | 0.3877 | 0.0553 |
| Jaccard Similarity | 0.2405 | 0.0433 |
| BLEU Score | 0.0898 | 0.0432 |
| ROUGE-L $F_{measure}$ | 0.2859 | 0.0474 |
| Levenshtein Ratio | 0.5118 | 0.0389 |
| Embedding Similarity | 0.8620 | 0.0399 |
| Precision (Tokens) | 0.3491 | 0.0760 |
| Recall (Tokens) | 0.4208 | 0.1100 |

These examples, provided in the appendix, highlight the model's varying performance across input strategies and offer insight into the impact of preprocessing on generation quality.

The suite of metrics captures three facets of quality:

1. **Lexical Overlap:** F1 Token Overlap ($\approx 0.39$) and Jaccard Similarity ($\approx 0.24$) indicate moderate word-level agreement. BLEU ($\approx 0.09$) and ROUGE-L ($\approx 0.29$) remain low, as they penalize missing or reordered n-grams in these relatively long texts.

2. **Character-Level Edit Distance:** The Levenshtein Ratio ($\approx 0.51$) reveals that only about half of all characters align in sequence, reflecting substantial paraphrasing.

3. **Semantic Fidelity:** A high Embedding Similarity ($\approx 0.86$) demonstrates that, despite surface-form differences, the model's outputs convey nearly the same meaning as the ground truth. Token-level Precision ($\approx 0.35$) vs. Recall ($\approx 0.42$) suggests the model includes extra or varied content (lower precision) while covering most ground-truth concepts (higher recall).

**Conclusion:** Although literal n-gram overlap is limited — resulting in low BLEU/ROUGE scores—the strong semantic correspondence suggests the core content is well captured. Future work should target entity-level fidelity and surface-form consistency (e.g., via constrained decoding or slot-filling) to boost lexical metrics without sacrificing meaning.

### API LLM evaluation

As shown in Table 2, *basic_outputs.jsonl* scored highest (4.05), *dp2_outputs.jsonl* was intermediate (3.91), and *dp1_outputs.jsonl* scored lowest (2.85). Scores were obtained via an API-driven LLM (DeepSeek V3 0324) prompted to rate each bulletin (1–10) on structure and content against RTF exemplars. This contrasts with earlier automated and structural analyses that favored the *dp1* format, suggesting human judges prize flexibility and contextual clarity over rigid templating. The large standard deviations further reveal variability in bulletin quality.

Interestingly, both fine-tuned variants achieved a somewhat lower score than the basic variant. This puts into question the effectiveness of the FT pipeline. However, finding the main culprit for this result is not very straightforward. Both data preprocessing and FT procedure include many parameters that can be futher tweaked. Additionally, the quantity, quality and variety of the data itself surely has some impact on the results.

**Table 2.** DeepSeek V3 API Evaluation Scores Summary

| Dataset | Average Score | Std. Dev. |
|---|---|---|
| basic_outputs.jsonl | 4.05 | 1.9049 |
| dp1_outputs.jsonl | 2.85 | 1.1522 |
| dp2_outputs.jsonl | 3.9130 | 1.8630 |

## Discussion

### Qualitative Analysis of Preparation Strategies

To better understand how data preparation impacts generation quality, we compared representative outputs from all three variants. DP1 yielded bulletins with high structural fidelity, closely mimicking RTV Slovenija formatting (e.g., *road* $\rightarrow$ *direction* $\rightarrow$ *event* $\rightarrow$ *consequence*). However, its rigid template often caused repetitive phrasing and excluded minor events outside the 8-hour window.

DP2 outputs, while lacking formal headers, showed more fluent and human-like summaries, often merging adjacent incidents effectively. These captured contextual nuance better, especially when events spanned multiple RTF paragraphs.

The basic variant produced mixed results: broader inclusion improved recall but occasionally introduced outdated or misaligned events due to the absence of temporal filtering.

Overall, LLM-based evaluations (e.g., DeepSeek V3) favored DP2 and basic over DP1, highlighting a trade-off between structural alignment (DP1) and natural summarization (DP2/basic). Future strategies may benefit from blending structural templates with adaptive, entity-aware inputs.

### Future Work

Future improvements could include integrating *named-entity alignment metrics* to better assess factual consistency (e.g., road names, locations). Adding *BERTScore* or other semantic-aware metrics would complement lexical evaluations.

On the data side, *adaptive temporal filtering* and hybrid strategies that combine DP1's structure with DP2's flexibility could improve both coverage and style.

Model-wise, exploring *constrained decoding* or *reinforcement tuning* (e.g., RLHF) may help enforce formatting and factual accuracy. Finally, incorporating *external data sources* such as weather or live traffic sensors could enable real-time applications.

## References

[1] ACL Anthology. Advances in automated journalism and natural language generation. *ACL Conference*, 2023.

[2] Arxiv.org. Transformer-based deep learning for natural language processing. *arXiv preprint*, 2022.

[3] Arxiv.org. Bias and objectivity challenges in ai-generated news. *arXiv preprint*, 2023.

[4] Tom Brown and et al. Language models are few-shot learners. *NeurIPS*, 2020.

[5] Teven Scao and et al. Bloom: A 176b parameter open-access multilingual language model. *arXiv preprint*, 2022.

[6] Linting Xue and et al. mt5: A massively multilingual pre-trained text-to-text transformer. *Journal of Machine Learning Research*, 2021.

[7] Matej Ulčar and Marko Robnik-Šikonja. Sloberta: The slovenian roberta model. *arXiv preprint*, 2021.

[8] Matej Ulčar and Marko Robnik-Šikonja. Slot5: Pre-trained text-to-text models for slovenian. *arXiv preprint*, 2022.

[9] Y. Zhang and et al. Fine-tuning strategies for large language models. *Journal of Artificial Intelligence Research*, 2022.

[10] L. Reynolds and K. McDonell. Prompt programming for large language models. *arXiv preprint*, 2021.

[11] Sebastian Schelter and et al. Automated event-driven news generation. *Data Mining and Knowledge Discovery*, 2019.

**Appendix: Prompt–Response Format Example**

```
VHOD:
<p><strong>Zastoji</strong></p>
<p>- na ljubljanski južni obvoznici od Rudnika proti razcepu Kozarje;
    <br>- na gorenjski avtocesti od Kosez do predora Šentvid proti Kranju;
    <br>- na štajerski avtocesti med Blagovico in Trojanami proti Celju ter pred
     ↪ prehodom Šentilj proti Avstriji;
    <br>- na ljubljanski zahodni obvoznici od Kosez proti Kozarjam ter na primorski
     ↪ avtocesti med Ljubljano in Brezovico proti Kopru;
    <br>- na cestah Ljubljana – Brezovica, Lesce – Bled in Šmarje – Koper.
</p>

IZHOD:
Zastoji so na cestah Šmarje-Koper in Lesce-Bled.
Prav tako so zastoji na zahodni in južni ljubljanski obvoznici od Kosez in Rudnika
 ↪ proti Primorski. Potovalni čas se na obeh odsekih podaljša za približno 15 minut.
Zaradi del je krajši zastoj na štajerski avtocesti med Blagovico in Trojanami proti
 ↪ Mariboru. Opozarjamo na nevarnost naleta.
```

**Figure 1.** An example of LLM output generation for given input data using only prompt engineering methods. The input (VHOD) is a flattened and HTML-tagged paragraph derived from structured traffic events, while the output (IZHOD) is a free-form textual summary in the style of RTV Slovenija traffic reports.

### Appendix: Basic Prompt–Response Format Example

```
VHOD:
Na štajerski avtocesti med priključkoma Slovenska Bistrica sever in Fram promet poteka
 ↪  po dveh zoženih pasovih v obe smeri.
Cesta Litija – Zagorje bo zaprta pri Šklendrovcu v soboto, 18. maja, med 7.30 in 17.
 ↪  uro.
Več o delovnih zaporah v prometni napovedi.
Oviran promet na gorenjski avtocesti med predorom Šentvid in Podutikom proti Kosezam.
Regionalna cesta Sevnica – Planina pri Sevnici je zaprta pri Orešju nad Sevnico.
Okvara vozila na primorski avtocesti pred razcepom Gabrk proti Kopru.
Prireditve ta konec tedna.
Zaradi poplavljenega vozišča je zaprta cesta Lešje – Koritno v občini Majšperk.
Povečan promet na cestah iz smeri Avstrije proti Hrvaški. Zastoji:- na gorenjski
 ↪  avtocesti prei priključku Jesenice vzhod proti Ljubljani;- upočasnjen promet na
 ↪  primorski avtocesti od Kozarij proti Brezovici;- na glavni cesti Postojna – Pivka,
 ↪  med Matenjo vasjo in Prestrankom.


IZHOD:

Na štajerski avtocesti med Slovensko Bistrico in Framom promet poteka po dveh zoženih
 ↪  pasovih v obe smeri.
Cesta Litija – Zagorje bo zaprta pri Šklendrovcu v soboto, 18. maja, med 7.30 in 17.
 ↪  uro.
Okvara vozila ovira promet na primorski avtocesti pred razcepom Gabrk proti Kopru.
Oviran promet na gorenjski avtocesti med predorom Šentvid in Podutikom proti Kosezam.
Regionalna cesta Sevnica – Planina pri Sevnici je zaprta pri Orešju nad Sevnico.
Na cestah iz smeri Avstrije proti Hrvaški je povečan prometni tok. Zastoji so na
 ↪  gorenjski avtocesti pred priključkom Jesenice vzhod proti Ljubljani, upočasnjen
 ↪  promet na primorski avtocesi med Kozarijami in Brezovico, ter na glavni cesti med
 ↪  Matenjo vasjo in Prestrankom.
```

**Figure 2.** Basic example.

**Appendix: DP1 Prompt–Response Format Example**

```
VHOD:

Prometne informacije       18. 08. 2024   \t   01.04          2. program

Podatki o prometu.

Zastoji Gorenjska: - Na gorenjski avtocesti med Lescam in predorom Karavanke proti
 ↪  Avstriji, 10 km. Izvoz Lipce je dovoljen samo za lokalni promet. - Na cestah
 ↪  Jesenice - Hrušica in Podkoren - Korensko sedlo. Delo na cesti Cesta Senarska -
 ↪  Lenart , pri Sveti Trojici bo zaprta do nedelje,18. avgusta, do 20. ure. Obvoz bo
 ↪  potekal po državnih cestah in po avtocesti med priključkoma Lenart { Sv. Trojica in
 ↪  obratno, tudi za vozila brez vinjete vendar samo med tema dvema priključkoma. Več o
 ↪  delovnih zaporah v prometni napovedi . Buy vignette for Slovenia online Thererore
 ↪  long queues are expected in entering points from Austria to Slovenia, i.e.
 ↪  Karavanke tunnel (A2) and Sentilj/Spielfeld crossing (A1). Important reason for
 ↪  these queues is that drivers don't have vignette for Slovenian roads and have to
 ↪  buy them at the border. To reduce or even avoid long waiting periods drivers are
 ↪  strongly recommended to buy vignette for Slovenian motorways online. They can do it
 ↪  here . DARS Janko Poženel, GNC Prireditve  Tovorni promet   Italija

IZHOD:
Prometne informacije        18. 08. 2024       01.04          2. program

Podatki o prometu.

Na primorski avtocesti je zastoj na izvozu Brezovica proti Kopru, čas vožnje se tam
 ↪  podaljša za približno 10 minut.

Na gorenjski avtocesti je zastoj pred predorom Karavanke proti Avstriji, čas potovanja
 ↪  se tam podaljša za približno pol ure.

Na glavni cesti Maribor-Hoče so zastoji v obe smeri med priključkom Maribor-vzhod in
 ↪  krožiščem Pobrežje. V smeri proti Mariboru je promet oviran na vipavski hitri cesti
 ↪  med priključkoma Vipava in Selo, zaradi pokvarjenega vozila.

Na mejnih prehodih Gruškovje in Obrežje vozniki tovornih vozil na vstop v državo čakajo
 ↪  2 uri, na Obrežju pa tudi na izstop.


GROUND TRUTH:
Prometne informacije      18.08.2024              05.00          1. in 2. program

Podatki o prometu.

Na štajerski avtocesti je zaradi prometne nesreče pred priključkom Fram proti Ljubljani
 ↪  zaprt vozni pas.

Na pomurski avtocesti je na izvozu Cerkvenjak proti Mariboru zaradi pokvarjenega vozila
 ↪  oviran promet.

Na gorenjski avtocesti je med Lescam in predorom Karavanke proti Avstriji približno 2
 ↪  kilometre dolg zastoj. Izvoz Lipce je dovoljen samo za lokalni promet.

Taradi prireditve bo v Ljubljani od 8-ih do 15-ih zaprta Štajerska cesta, med rondojem
 ↪  Tomačevo in rondojem s Kranjčevo cesto.

Cesta Senarska - Lenart pa bo zaradi del pri Sveti Trojici zaprta do 20-tih. Obvoz je
 ↪  po državnih cestah in po avtocesti med priključkoma Lenart - Sv. Trojica in
 ↪  obratno, tudi za vozila brez vinjete vendar samo med tema dvema priključkoma.
```

**Figure 3.** Example of output from GaMS-27B-Instruct using the DP1 method for prompt construction and fine-tuning. The corresponding ground truth summary is shown for comparison.

**Appendix: DP2 Prompt–Response Format Example**

```
VHOD:

Danes ponoči bo skozi predor Karavanke od 20. do 5. ure, promet potekal izmenično s
 ↪  čakanjem pred predorom predvidoma 30 minut.
Na ljubljanski severni obvoznici potekajo dela med Zadobrovo in Tomačevim. Med Novimi
 ↪  Jaršami in Zadobrovo promet proti Zaloški poteka po dveh zoženih pasovih.
Čakalna doba: Obrežje in Gruškovje.
Na štajerski avtocesti med Polskavo in Framom proti Mariboru zaprt vozni pas. Nastaja
 ↪  zastoj, 1,8 kilometra.
Zaprte ceste zaradi gašenja požara na Primorskem:- regionalna cesta Črni Kal -
 ↪  Gabrovica pri Črnem Kalu - Osp - meja z Italijo;- lokalna cesta med Črnim Kalom in
 ↪  Staro Gabrovico;- ter lokalna cesta na relaciji avtocestni priključek Kastelec -
 ↪  vas Kastelec - Socerb - meja z Italijo.
Na cesti Bohinjska Bela - Bled, na Bledu.

IZHOD:

Na gorenjski avtocesti proti Ljubljani je zaradi nesreče zaprt prehitevalni pas med
 ↪  priključkoma Brezje in Podtabor. Nastal je zastoj, opozarjamo na nevarnost naleta.
 ↪  Proti Karavankam pa je zaradi pokvarjenega vozila zaprt vozni pas pred predorom
 ↪  Ljubno. Zaradi del bo na gorenjski avtocesti med 11-to in 16-to uro promet skozi
 ↪  predor Karavanke potekal izmenično enosmerno. Na mejnem prehodu Obrežje vozniki
 ↪  tovornih vozil na vstop v državo čakajo 2 uri, v Gruškovju pa do pol ure. Zaradi
 ↪  gašenja požara so zaprte regionalna cesta Črni Kal-Gabrovica pri Črnem
 ↪  Kalu-Osp-meja z Italijo, lokalna cesta Črni Kal-Stara Gabrovica ter lokalna cesta
 ↪  od avtocestnega priključka Kastelec, skozi Socerb, do meje z Italijo. Zastoj je na
 ↪  cesti Bohinjska Bela-Bled.
```

**Figure 4.** Example of output from GaMS-27B-Instruct using the DP2 method for prompt construction and fine-tuning.

**Appendix: Manual ChatGPT Comparative Evaluation of Output Sets**

```
Output Set        Score
-------------------------
basic_outputs     4 / 10
dp1_outputs       9 / 10
dp2_outputs       6 / 10

Rationale:
1. basic_outputs (4/10)
   - No standard header (\Prometne informacije ...").
   - Inconsistent ordering of road, direction, event.
   - Missing \Podatki o prometu." section title.
   - Uneven naming conventions and event grouping.

2. dp1_outputs (9/10)
   - Correct header and date/time/program line.
   - Includes \Podatki o prometu." before events.
   - Consistent sentence structure:
       road + direction → event → consequence.
   - Proper motorway names and Slovene terminology.
   - Minor slip-ups in program numbering/order.

3. dp2_outputs (6/10)
   - Well-formed event sentences (location, reason, impact).
   - Omits the bulletin framing (header + title).
   - Lacks blank lines separating items.
   - Presents raw list rather than full RTF-style bulletin.
```

**Figure 5.** Scores and qualitative rationale for each output set, based on the formatting rules in `PROMET.docx` and style exemplars in the provided RTF files by ChatGPT.