SLaLoM 2026

2nd Slovenian workshop on Large Language Models Techniques and Applications & EMMA project meeting

Kranjska Gora, Slovenia, February 12th-13th, 2026

Workshop proceedings

Edited by Jaya Caporusso and Nada Lavrač

Jožef Stefan Institute, Ljubljana, Slovenia

Workshop Agenda

Thursday, February 12, 2026

13:00 - 13:30 Gathering and registration

13:30 - 14:00 Welcome address, overview of EMMA goals and main outcomes (Nada Lavrač)

14:00 - 14:30 State of the art of AI in the media industry and relevance of EMMA achievements to Kliping (Sebastjan Hribar)

14.30 - 15.30 Presentation of EMMA results

15.30 - 16.00: Coffee break

16:00 - 18.00: Large Language Models Techniques and Applications - focus on news industry

18:30 - 19:00 Summary and discussion

18:30 - 19:30 Dinner

Friday, February 13, 2026

9:00 - 9:30 Gathering

9:30 - 12:30 Summary of achievements and future plans

12:30 - 13:30 Lunch

13:30 - 16:30 Informal discussions and activities

Boshko Koloski: Keywords extraction and targeted keyword extraction

This study detailed the development of advanced methods for keyword extraction in multilingual and cross-lingual environments, addressing challenges related to diverse textual data and limited training resources. We introduced SEKE (Specialised Experts for Keyword Extraction), a supervised approach based on the Mixture of Experts (MoE) framework with DeBERTa as the backbone, enhanced by a BiLSTM layer to enable effective learning even on smaller corpora. The routing mechanism allows experts to specialise in different linguistic features, improving both performance and interpretability.

Extensive benchmarking across multiple English datasets demonstrated state-of-the-art performance compared to strong supervised and unsupervised baselines. Analysis of expert behaviour revealed that different experts naturally specialise in distinct syntactic and semantic patterns, such as punctuation, stopwords, parts-of-speech, or named entities, offering insights into the model's internal decision processes.

In addition, we evaluated supervised and unsupervised zero-shot cross-lingual keyword extraction methods in multilingual settings. Results show that embedding-based approaches outperform traditional statistical and graph-based methods, with zero-shot transfer between languages consistently surpassing unsupervised techniques. However, experiments also revealed that performance declines as more languages are introduced into the training set, highlighting important trade-offs in multilingual model design.

Nikola Ljubešić: Recursive news classification into the IPTC hierarchy

Automatic classification of news articles according to the IPTC Media Topic NewsCodes schema is a practically important but challenging task, particularly when fine-grained categorization across multiple levels of the hierarchy is required. While fine-tuned encoder-based transformer models have proven effective at the top level of the hierarchy – as demonstrated by our multilingual XLM-R-based classifier covering 17 top-level IPTC categories across all XLM-R-supported languages – extending this approach to deeper hierarchical levels is severely hindered by data scarcity, as labeled instances for specific lower-level categories are difficult to obtain at scale. In our contribution, we propose a recursive classification framework that addresses this limitation by shifting from instance-based to definition-based classification for sub-levels of the hierarchy. Specifically, we employ decoder-based large language models conditioned on the textual definitions of child categories rather than on labeled training examples, enabling zero-shot classification at each hierarchical level. The process is applied recursively, layer by layer, starting from the top-level prediction produced by the fine-tuned encoder model. Users can control classification depth by specifying both the input text and the portion of the IPTC schema to traverse, offering a flexible and scalable solution for hierarchical news topic classification.

Nikola Ivačič: News categorisation into tailored topics of Kliping

In this work, we investigated extreme multi-label text classification (XMC) in real-world industrial environments, with a particular focus on less-represented languages and low-resource settings, such as Slovene media monitoring. XMC requires assigning documents to multiple labels from extremely large label sets that may contain thousands of categories. In practical deployments, systems must process large volumes of text efficiently while adapting to evolving label spaces and limited computational resources. We therefore analyzed the limitations of large transformer models and other resource-intensive approaches for such production environments.

We evaluated several approaches on two datasets: $NewsMon_{sl}$, a Slovene media monitoring dataset, and EURLEX57K, a widely used benchmark for extreme multi-label classification. The tested methods included TF-IDF, one-vs-all logistic regression and SVM, ML-KNN, XLM-RoBERTa-base, and dense retrieval models such as BGE-M3 and its hard-negative fine-tuned variant. These models allowed us to compare traditional machine learning baselines, transformer-based encoders, and modern embedding-based retrieval approaches across different data conditions and label distributions.

Building on the RAE-XMC (Retrieval-Augmented Extreme Multi-Label Classification) framework, we implemented a simplified RAE-XMC pipeline adapted to scenarios where label descriptions are unavailable. Instead of retrieving both label descriptions and training examples, our approach retrieved the most similar training documents, transformed similarity scores into temperature-scaled weights, and aggregated their label vectors to generate predictions. Our experiments showed that retrieval-based approaches can remain competitive while requiring significantly fewer computational resources, making them a practical solution for large, dynamic label spaces and low-resource multilingual deployments.

Nishan Chatterjee: Target-based sentiment analysis

We wrote a journal paper based on our findings which contributes a few key things to the field, and has a strong focus on practical applications for South Slavic languages. Our main contributions include the AspectBench Benchmark, a new, large-scale, and publicly available subset for document-level ABSA in Serbo-Croatian, along with benchmarking on Slovenian. We built this from real-world news media to reflect actual industrial challenges, focusing specifically on how models generalize between seen and unseen aspects. We also conducted the first systematic empirical study for South Slavic news, where we looked closely at performance versus complexity trade-offs by comparing lightweight embedding classifiers against fine-tuned Pre-trained Language Models (PLMs), Hierarchical Attention Networks (HANs), and quantized Large Language Models (LLMs). Additionally, we developed a novel method for practical and explainable LLM application. By creating a hybrid approach where a specialized model's predictions guide an LLM's Chain-of-Thought reasoning using DSPy, we made local, open-source LLMs competitive in specialized tasks while leaving a faithful, auditable reasoning trail that bridges the gap between accuracy and interpretability.

Regarding the current status and next steps, we initially submitted the paper to PeerJ and the Journal of Intelligent Information Systems. It was desk-rejected from both, primarily because PeerJ requires all data to be open, but part of ours is under NDA and while the other part is proprietary; and JIIS deemed it to be out of scope. We have since resubmitted it to Applied Intelligence. The paper ("Evaluating Fine-Tuned, Embedding-Based, and Zero-Shot Models for Aspect-Based Sentiment Analysis in South Slavic News," tracking number APIN-D-26-01921) has been "With the Editor" since February 18, 2026. For my next tasks, based on our discussions at the Kranjska Gora meeting, I need to finalize the release pipeline. The data preparation is already done, so now I need to clean up the code and prepare a release file for HuggingFace so we can get the models out there. Once everything is cleaned up, I will share it with the rest of the team for review before changing the visibility to public.

Boshko Koloski, Senja Pollak: Sentiment analysis for ESG goals

Environmental, Social, and Governance (ESG) considerations play an increasingly important role in evaluating corporate performance, reputation, and long-term sustainability. However, reliable ESG ratings remain scarce for smaller companies and emerging markets, particularly in low-resource languages such as Slovene. To address this gap, we introduce the first publicly available Slovene ESG sentiment dataset together with a set of models for automatic ESG sentiment detection.

The dataset is constructed from the MaCoCu Slovene news corpus, combining large language model (LLM)–assisted filtering with human annotation of company-related ESG content. We evaluate multiple modelling approaches, including monolingual and multilingual transformer classifiers (SloBERTa, XLM-R), embedding-based models (TabPFN), hierarchical ensemble architectures, and large language models.

Results show that LLMs achieve the strongest performance on Environmental (Gemma3-27B, F1-macro: 0.61) and Social aspects (gpt-oss-20B, F1-macro: 0.45), while fine-tuned SloBERTa performs best on Governance classification (F1-macro: 0.54). Finally, we demonstrate the practical utility of the approach through a case study applying the best-performing model (gpt-oss) to analyse ESG coverage of selected companies in Slovene news over time.

Marko Pranjić: Longitudinal media monitoring

This presentation detailed the development of a high-performance system for Longitudinal Media Monitoring, designed to identify and track semantic shifts across extensive news media timelines. While traditional semantic change detection methods often struggle with computational overhead, this work addresses the scalability bottleneck inherent in state-of-the-art approaches. By applying a refined methodology to the sPeriodika dataset (spanning 150 years of Slovene periodicals) we demonstrated that the proposed system is

highly scalable, offering significantly deeper insights into word temporal dynamics than previous models.

Beyond performance metrics, the session highlighted critical theoretical improvements regarding model training. Specifically, we identified a significant weakness in the AnglE loss function, noting that it lacks symmetry. To rectify this, we presented a novel solution that simultaneously resolves the cosine saturation problem and the symmetry issues introduced by AnglE. Supported by gradient and function value plots, the results show a more stable and theoretically sound approach to training. Next steps will focus on fine-tuning the model for word sense disambiguation (WSD) to further refine the sensitivity of detected semantic trends.

Aleš Žagar: Story identification and document summarisation

The discussion focused on consolidating progress and defining next steps. The paper on story identification will be resubmitted. We also discussed the development of a new human-annotated summarization dataset that will support both general summarization and target-based (query-focused) summarization. Several constraints were identified in relation to advancing text summarization research: (i) hardware limitation; (ii) methodological challenges in using approximately 1,000 annotated samples; and (iii) scalability requirements, since the intended use case involves processing between ten and twenty thousand articles per day, imposing strict efficiency and throughput constraints. The proposed target-based summarization approach, QFS-Composer, will be evaluated on the new dataset. This framework targets query-focused summarization in Slovene and integrates query decomposition, question generation (QG), question answering (QA), and abstractive summarization.