

DATA MINING AND DATA WAREHOUSES – SIKDD

ODKRIVANJE ZNANJA IN PODATKOVNA SKLADIŠČA - SIKDD

Wednesday 7th October 2026 / Sreda 7. oktober 2026

Physical Seminar Room / Fizikalni Seminar

Link for online attendance / Povezava za spletno udeležbo:

<https://zoom.us/j/94038351112?pwd=VqaV86n8Mwmahqak7AgBNHUqphleXO.1>

9:00 – 11:00	Session A / Sekcija A - Chair / Vodi Dunja Mladenić • Domen Jeršek, Blaž Kažič and Klemen Kenda Exploring Feedback-Aware Retrieval for IT Support First-Reply Generation • Inna Novalija, Janez Brank, Dunja Mladenić and Marko Grobelnik The Changing Slovenian Smellscape, 1870–1919: Measuring Olfactory Discourse • Drejc Pesjak, Dunja Mladenić and Jan Rupnik Explaining Headline A/B Test Outcomes with Sparse Autoencoder Features • Georgi Trajkov, Blaž Škrlj, Marko Grobelnik and Dunja Mladenić Same Labels, Smaller Bill: An Ablation Study of LLM Claim Classification in Cross-Source News • Natalija Mečkaroska and Georgi Trajkov A Hybrid Lexicon-LLM Approach to Skill Extraction from Job Advertisements • Oleksandra Topal, Erik Calcina, Luciano Telesca and Dunja Mladenić Benchmarking Machine-Learning Forecasts of Low-Magnitude Seismicity in the Italy–Slovenia Border Region • Abdul Sittar, Erik Novak and Erik Calcina The Effect of Label Only vs. Label Plus Description: Comparing GLiNER1 and GLiNER2 for Clinical Entity Recognition Across Five Languages • Nina Kokalj, Erik Calcina, Abdul Sittar and Erik Novak Data Efficiency of GLiNER: A Multilingual and Multi-Scale Learning Curve Study
--------------	---

11:00 – 11:30	Coffee Break / Odmor za kavo
11:30 – 13:30	Session B / Sekcija B - Chair / Vodi Marko Grobelnik • Rok Klančič and Klemen Kenda Comparing Pseudo-Labeling Criteria for Active Learning on a Service-Desk Dataset • Georgi Trajkov LLM-Generated Song Lessons: Perceived Usefulness and Shared-Cache Cost Analysis • Adrian Mladenich Grobelnik Reliable but Not Right: Automated Evaluation of Generated R&D Ideas • Nadezhda Komarova and Inna Novalija Graphs of n-grams for Text Generation and Characteristic Phrase Extraction in the Domain of Smell • Vuk Đuranović and Marko Robnik-Šikonja QFS-Composer: Query-Focused Summarization Pipeline for Less Resourced Languages • Jakob Vokač, Marko Grobelnik and Dunja Mladenich Beyond Named Citations: A Corpus of Resolved Cross-References in EU Legislation • Andrej Natev, Elham Motamedi and Inna Novalija Towards Scientific Contradiction Mining: Topic Discovery and Claim Extraction from AI Literature • Žiga Klun, Dragi Kocev and Jure Brence Chronos-2 for Spacecraft Telemetry Forecasting • Nadiia Shpanko, Erik Calcina and Erik Novak Biomedical Relation Extraction: What Lies Between the Entities? • Kristina Piiarska, Aleksandar Tošić and Pau Brunet The Role of Artificial Intelligence in Early Moisture Risk Detection for Wooden Buildings
13:30 – 14:00	Lunch break / Odmor za kosilo

14:00 – 15:30	Session C / Sekcija C - Chair / Vodi Marko Grobelnik • Drejc Pesjak, Georgi Trajkov, Luis Rei and Dunja Mladeníć E3 Atlas: An Autonomous LLM Agent for Curating and Serving a Departmental Knowledge Base • Seyed Iman Hosseini, Lorenzo Seghesio, Konstantinos Gryllias and Dunja Mladeníć Evaluating Large Language Model-Generated Explanations for Defective-Part Prediction in Injection Moulding Using Feature-Attribution Evidence • Simona Cholakova, Domen Vake and Aleksandar Tošić Learning Molecular Representations from SMILES for Acute Toxicity Prediction • Yunfei Liang, Matic Fučka and Danijel Skočaj Logical Anomaly Detection Using Segmentation Transformers • Andrej Srakar Posterior Contraction Rate and Roadmap to Semiparametric Bernstein-von Mises Theorem for Bayesian Causal Forests • Jaka Čibej Validation of Proper Noun Lexemes in the Digital Dictionary Database of Slovene Using Wikidata and Large Language Models • Sidorela Uku and Klemen Kenda Comparative Analysis of Feature Selection and Imbalance Handling for Semiconductor Defect Detection on the SECOM Dataset • Gašper Leskovec, Costas Mylonas and Klemen Kenda Operationalising Active Learning for Power Grid Security Assessment: A Human-in-the-Loop Retraining Service for a Digital Twin Platform
15:30 – 16:00	Coffee Break / Odmor za kavo

16:00 – 17:15	Session D / Sekcija D - Chair / Vodi Dunja Mladenić • Aleksandar Brodschneider and Matjaž Gams From Quantum Machine Learning to Quantum Artificial Intelligence: A Systematic Meta-Review • Lovro Bobnar From Explainability to Accountability: XAI and Fundamental Rights Impact Assessments under the EU AI Act • Kaja Praprotnik, Niki Hrovatin, Jernej Vičič and Domen Vake Influencing Sentiment in LLM Output via Token Selection Manipulation • Dario Mammana and Petra Kralj Novak From Label Agreement to Downstream Performance: Evaluating LLM-Generated Training Data • Gasper Kreft, Aljaž Osojnik, Martin Žnidaršič and Bernard Ženko Short Term Time-Series Forecasting Framework • Gal Zmazek, Aljaž Osojnik, Martin Žnidaršič and Bernard Ženko Short-Term Load Forecasting at the Device Level: Are Specialized Deep-Learning Approaches Worth the Effort? • Joao Pita Costa, Andrej Srakar, Oleksandra Topal and Jan Šturm Data-Driven and Model-Based Causal Discovery in Sparse Multivariate Time Series: A Comparative Framework for Sustainable Development Systems • Stefan Kofilovski and Georgi Trajkov Legal Categories from Incident Reports: Evaluating 14 LLMs under the EU AI Act • Sivaramakrishnan R Guruvayur, Krishnasree Achuthan and Joao Pita Costa Balancing Sustainable Development Goals through Orchestrating Autonomous Multi-Agent Architecture • Martin Macak and Marko Grobelnik Accurate but Uninformative: Checking an LLM Assessment Pipeline Without Ground Truth • Luka Petek and Vili Podgorelec Combining Supervised and Unsupervised Learning on SMART Data for Hard Drive Failure Assessment • Martin Macak Advertising Is Not Noise: Sponsored Speech Outranks New Terminology in Automatic Term Extraction
---------------	--