AI 4 SCIENCE

LJUBLJANA, SLOVENIA

International Conference

AI for SCIENCE 2025

including DS-2025, the SMASHing conference (4 tracks)
and 4 additional tracks

22.09.2025 - 26.09.2025

University of Ljubljana,
Faculty of Computer and Information Science
Ljubljana, Slovenia

Contents

General

Organizers

University of Ljubljana, Faculty of Computer and Information Science
SMASH
The Jožef Stefan Institute
Slovenian Artificial Intelligence Society

Committee

Sašo Džeroski, Jožef Stefan Institute - General Chair, SMASH supervisor
Vida Groznik, Faculty of Computer and Information Science, University of Ljubljana - Local Organization Chair, SMASH supervisor
Gabrijela Zaharijas, University of Nova Gorica - Coordinator of SMASH tracks, SMASH supervisor and PI

Local Organization Committee

Vida Groznik, University of Ljubljana - Local Organization Chair, SMASH supervisor
Aleksander Sadikov, University of Ljubljana - SMASH supervisor
Maj Zirkelbach, University of Ljubljana
Mila Marinković, University of Ljubljana
Sašo Džeroski, Jožef Stefan Institute - SMASH supervisor
Sintija Stevanoska, Jožef Stefan Institute
Laura Šušteršič Zorn, University of Nova Gorica
Ana Čufer, Slovenian Artificial Intelligence Society

Webmasters

Maj Zirkelbach, University of Ljubljana
Tilen Gabršček, University of Ljubljana
Simon Bele, University of Ljubljana

Venue

Main venue:

AI 4 SCIENCE 2025 will take place between 22 and 26 September 2025 at the University of Ljubljana's Faculty of Computer and Information Science (UL FRI), located at **Večna pot 113, Ljubljana, Slovenia**.

The faculty (UL FRI) enrolls about 1,600 students. It has just recently (2014/2015) built its own modern building. It offers a wide variety of air conditioned classrooms, all equipped with state-of-the-art AV equipment, including at least two overhead projectors, wireless internet (eduroam), whiteboards, and plenty of electricity sockets for notebook computers. The faculty has two large lobbies where the participants can mingle and network, interspersed with a plenty of small cosy workspaces equipped with electricity, and a spacious terrace with benches just outside the main entrance.

Social events:

The **welcome reception** will be held in front of the school's main venue on its big open-air platform. The location is ideal for hanging out in a relaxed atmosphere in the summer evenings from where the attendees can observe the colorful sunset. We will offer fingerfood to present local couisine along with local wines and non-alcoholic beverages.

The **social dinner** will be held at the Hotel Union, located on **Miklošičeva cesta 1**, in the centre of Ljubljana.

Poster sessions

Meals

Internet Access

Code of Conduct

We strive for making every attendee feel welcome and respected. Do not hesitate to contact any of the organizers if you are the witness or the victim of any discrimination or harassment. Step up and speak out to stop any kind of inappropriate behaviour you witness. Bystander intervention creates safer communities and prevents harmful escalation.

Other

The Emergency number for Slovenia is 112.

Sponsors

We acknowledge the generous support of the Faculty of Computer and Information Science of the University of Ljubljana, SMASH, Jožef Stefan Institute, Ljubljana Tourism, Ljubljanski potniški promet, and Project - Fondazione FAIR.

Program

AI & Digital Humanities

Track Chairs

Marko Robnik Šikonja, University of Ljubljana - SMASH supervisor
Špela Arhar Holdt, University of Ljubljana
Senja Pollak, Jožef Stefan Institute

09:00 - 09:50	**Marieke van Erp** Reflections on Reflection in Humanities and AI
09:50 - 10:10	AI for Managing Complex Textual Traditions: How to Handle Multiple Variant Readings?
10:10 - 10:30	Simulating Archival Biases for Interpretable Surrogate Scholar Learning
10:30 - 11:00	**Coffee Break**
11:00 - 11:20	Copyright obstacles in building large language model in Slovenia: exceptions in the copyright law, right clearance via licenses and obstacles in practice
11:20 - 11:40	Using large language models to generate distractors for language games
11:40 - 12:12	SloPragEval: Creating the First Pragmatics Understanding Benchmark for Slovene
12:00 - 12:20	Evaluating LLMs on Value Annotation Task
12:20 - 12:40	Agentic Large Language Models for Grammatical Analysis of Multilingual Corpora
12:40 - 13:00	Attempt to Create Synthetic Dataset for Grammar Error Correction in Slovenian Language
13:00 - 14:30	**Lunch Break**
14:30 - 15:20	**Federico Pianzola** AI for the Computational Study of Narrative and Readers
15:20 - 15:40	SloLitAA: Slovenian Literary Authorship Attribution via AutoML
15:40 - 16:00	Automatic detection of folkloristic motifs with large language models: the Cinderella tale
16:00 - 16:30	**Coffee Break**
16:30 - 17:20	**Slavko Žitnik** From Datasets to Data Spaces: What they are and what to expect?
17:20 - 17:40	Teaching Vision-Language Models Semiotics: Toward a Socio-Semiotic Framework for Multimodal AI

Reflections on Reflection in Humanities and AI

Marieke van Erp[1]

[1] *KNAW Humanities Cluster, University of Amsterdam*

The fields of Computer Science and Humanities operate in different ways: different research methods, different publication cultures, and different perspectives. But with interdisciplinary collaborations, projects, and teams, the use of computational methods is becoming more widespread in humanities use cases. As many archives have been and are still being digitised, automatic analysis methods are a necessity. However, historical texts present computational analyses with many different challenges such as digitisation artefacts, segmentation, language evolution, and changing societal values and large generic models are often not optimised for this domain resulting in suboptimal performance. In this talk, I will argue that combining the skills of humanities scholars with AI is beneficial to both fields: humanities scholars can benefit from new tooling to deal with big data in their domain, and AI researchers can benefit from the humanities' long tradition of source and tool criticism.

AI for Managing Complex Textual Traditions: How to Handle Multiple Variant Readings?

Serena Crespi[1], Alice Gydé[1], Carlos-Emiliano González-Gallardo[1], Elena Pierazzo[1]

[1] *Université de Tours*

In our presentation, we explore how artificial intelligence (AI) can be integrated into digital scholarly editing, especially for modern manuscript traditions. While XML-based standards have enabled structured encoding and publication, editorial workflows remain mostly manual and labour-intensive. We propose a workflow that incorporates AI at multiple stages—from image preprocessing and handwritten text recognition (HTR) to analyzing textual variation. To address limitations in training data, we employ multimodal data augmentation to generate synthetic handwriting samples that reflect the visual and material complexity of early modern manuscripts. This improves model performance without relying solely on manual transcription. Additionally, we employ unsupervised clustering techniques to identify structural patterns and differences across textual witnesses, utilizing visual tools such as dendrograms and heatmaps to enhance interpretability. Rather than replacing scholarly expertise, we see AI as a means to extend it, enabling handling of editorial tasks at scale while maintaining critical oversight. Our approach advocates for transparent, reflective use of computational tools that respect both the uniqueness of historical sources and the intearpretive work of humanists.

Simulating Archival Biases for Interpretable Surrogate Scholar Learning

Ricardo Inácio[1], Zafeiris Kokkinogenis[1], Vitor Cerqueira[1], Carlos Soares[1]

[1] *Faculdade de Engenharia da Universidade do Porto*

Agent-based models (ABM) foster a computational approach to study the humanities, by employing historical archives as foundations for their simulations. For example, in the work of Roman, encounters of Spanish explorers with indigenous Pueblos are represented through simple actions, and the emergence of complex cultural identities that align with the actual landscape of the Borderlands validated the method as a reliable means of reconstructing the depicted events. Nonetheless, it is plausible that such archives are biased towards those who recorded and preserved them. Namely, a selective inclusion of events

by the Spanish could lead to the omission of Indigenous voices. We aim to use an ABM to produce synthetic chronicles while injecting biases at different levels. Afterwards, these are fed into a machine learning classifier that acts as a surrogate scholar, to simulate the process of learning from the records. Then, the predicted probabilities allow us to interpret how the presence of these biases affects behavioural perceptions, by assessing proficiency through metrics disaggregated by culture. We observed that the injected biases effectively skew the perception of the surrogate in favour of the Spanish. Shapley interaction values also highlight that biases have different effects on assessments based on the confidence of the surrogate.

Copyright obstacles in building large language model in Slovenia: exceptions in the copyright law, right clearance via licenses and obstacles in practice

Maja Bogataj Jancic[1], Simon Krek[2]

[1] *ODIPI*
[2] *Jožef Stefan Institute*

"Is it necessary to obtain a licence from rightsholders in order to train "machines" on copyright works or is the training allowed on the basis of copyright exceptions? " Lawmakers around the world have approached this issue in different ways. The EU legislator introduced two new exceptions for text and data mining (TDM): an exception the purposes of scientific research and a "general" exception for all other purposes, including commercial use. These two exceptions were transposed into Slovenian law through an amendment to the Copyright and Related Rights Act. Article 57.b governs the exception for TDM for scientific research purposes, while Article 57.a ZASP applies to all other purposes, including commercial ones. Both exceptions allow the free reproduction of lawfully accessed works for the purpose of TDM. They also permit the digitisation of analogue content and remote access to analogue content for TDM purposes. In the case of the scientific research exception, the sharing and making available to the public of the results of TDM is also permitted. One of the biggest challenges is the definition of "lawful access" under ZASP because the term does not explicitly refer to access to content that is freely available online. Slovenian legal order has a special Legal Deposit Act which provides basis for lawful access to the "treasures" kept by National Library. Accordig to this law institutions have a duty – not merely an option – to ensure (lawful) access to deposited publications for researchers and research institutions when used for research purposes. Slovenian legal order provides a very solid copyright framework for building LLMs, but the institutions in charge still hesitate to exploit them in full. The presentation will present the legal basis and outline the challenges in practice. Regardless of the options provided by law the LLM builders in Slovenia try to clear rights with rightholders also via licenses. The presentation will present advantages of this right clearance model and challenges that it brings in practice. The presentation will also address a vider EU challenge of LLM builders, specifically open source LLM builders, in the field of copyright.

Using large language models to generate distractors for language games

Tjaša Arčon[1], Iztok Kosem[2], Špela Arhar Holdt[2]

[1] *University of Ljubljana, Faculty of Computer and Information Science*
[2] *University of Ljubljana, Faculty of Arts*

Word games support vocabulary acquisition, engagement, and cognitive development. CJVT Igre is a new digital platform featuring various Slovene word games aimed at a broad audience, from students to language enthusiasts. A major challenge in game development lies in preparing high-quality data. The games draw primarily on the Digital Dictionary Database of Slovene, however, even with curated lexical resources, additional annotation, filtering and manual oversight are often required to ensure the quality of the game content. We focus on two games requiring distractor generation (incorrect words presented alongside correct one) for synonym selection and collocation recognition. We explored how large language models (LLMs), specifically GPT-4o and Gemini-2.0-flash, can assist with preparing such data. Using a structured prompt, we tested both models on a sample of 30 Slovene synonym pairs. Gemini-2.0-flash outperformed GPT-4o in generating frequent, part-of-speech-consistent distractors while avoiding semantic overlap. Subsequently, it was used to generate over 50,000 distractors. Preliminary evaluation indicates fewer than 0.4% of outputs were true synonyms, while around 10% were likely invalid Slovene words. Manual review confirms most distractors meet the defined constraints, with remaining issues typically involving part-of-speech mismatches. We also examined distractor difficulty using similarity metrics such as Levenshtein distance. At the conference, we will present the synonym task methodology and results, ongoing work on collocation distractors, and reflections on the use of LLMs for language game content.

SloPragEval: Creating the First Pragmatics Understanding Benchmark for Slovene

Mojca Brglez[1], Špela Vintar[1]

[1] *Faculty of Arts*

Large language models are demonstrating increasing capabilities, excelling at benchmarks once considered very difficult. As their capabilities grow, there is a need for more challenging evaluations that go beyond surface-level linguistic competence. Namely, language competence involves not only syntax and semantics but also pragmatics, i.e., understanding situational meaning as shaped by context as well as linguistic and cultural norms. Recent studies have begun probing LLMs on such nuanced language, including figurative language and conversational implicatures. To contribute to this line of research, we introduce SloPragEval, the first pragmatics understanding benchmark for Slovene. It contains 300 scenarios with utterances that potentially flout Gricean conversational maxims, where the task is to infer the intended meaning of an utterance in a multiple-choice question answering setting. We discuss the difficulties of translation, describe our crowdsourcing campaign to establish a human baseline, and report pilot evaluations with open-source LLMs. Our results indicate that current models still struggle to infer implied speaker meaning in non-literal utterances. Finally, we argue that benchmarks targeting nuanced language understanding must be designed with care, constructed and validated with human responses, as opposed to comparing performance only between language models.

Evaluating LLMs on Value Annotation Task

Ajda Pretnar Žagar[1]

[1] *Faculty of Computer and Information Science, University of Ljubljana*

Human and civic values are foundational to public discourse and policy. Automating value annotation could support large-scale discourse analysis, policy alignment, and cross-cultural comparisons. However, current approaches suffer from limited value taxonomies, relying primarily on Schwartz's theory of basic values, and under-explored performance of large language models (LLMs). We present a multilingual value list refined by human experts

and evaluate its use in manual and automated annotation across seven languages (Danish, Dutch, English, Finnish, French, Swedish, and Slovenian). Our findings show both low inter-annotator agreement and inconsistent model performance, underscoring the subjective nature of value annotation. We argue that current models may support only shallow analysis and that meaningful improvement depends on better instruction-following, contextualization, and value list design.

Agentic Large Language Models for Grammatical Analysis of Multilingual Corpora

Matej Klemen[1], Kaja Dobrovoljc[2], Luka Terčon[2], Nives Hüll[1], Tjaša Arčon[1], Marko Robnik-Šikonja[1]

[1] *University of Ljubljana, Faculty of Computer and Information Science*
[2] *University of Ljubljana, Faculty of Arts*

Our work explores how large language models (LLMs) augmented with Universal Dependencies (UD) data can simplify and enhance multilingual grammatical analysis. We present two phases of the development of our system: (1) a baseline evaluation of zero-shot LLMs to assess their intrinsic linguistic knowledge, and (2) an agentic retrieval-augmented generation (RAG) system that integrates UD data. In the latter, LLMs generate code to analyze syntactically annotated UD instances, filtering out non-relevant data and synthesizing answers to user queries. Our evaluation, based on 17 selected grammatical features from the World Atlas of Language Structures (WALS), is conducted across languages with overlapping UD and WALS coverage. We assess system performance using both WALS labels and statistics extracted from UD data as ground truth. Initial results show that larger models like GPT-4o and LLaMA-3-70B outperform smaller models but still exhibit limitations. This work-in-progress points toward the potential and limitations of LLMs in corpus-informed grammar research and outlines future improvements, including demonstration-based explanations and tool calling.

Attempt to Create Synthetic Dataset for Grammar Error Correction in Slovenian Language

Gašper Jelovčan[1], Marko Robnik-Šikonja[1], Špela Arhar Holdt[1], Domen Vreš[1]

[1] *University of Ljubljana, Faculty of Computer and Information Science*

This research paper addresses the gap in the availability of grammar error correction datasets for the Slovenian language, a less-resourced language in the field of natural language processing. The primary objective is to create a comprehensive training dataset that incorporates various types of grammatical errors, sourced from both authentic and manually crafted examples. The methodology includes the collection of grammatically correct Slovenian texts, systematic error insertion, and manual verification to ensure authenticity and accuracy. A seed dataset containing 7,386 sentences with annotated grammatical errors has been developed, which will be expanded through the fine-tuning of a Large Language Model to generate approximately five million examples reflecting realistic language use. The final dataset will be used to fine-tune the GaMS Large Language Model for effective grammar error correction. This study not only highlights the challenges faced in developing language resources for less-resourced languages but also presents a novel methodology that can be adapted for similar languages, aiming to facilitate advancements in language correction tools and educational applications.

AI for the Computational Study of Narrative and Readers

Federico Pianzola[1]

[1] *University of Groningen*

Narrative can be conceptualised as a complex system, with pragmatic benefits for our description of interpretative and affective processes involved in reading. In this talk, I show how to use AI tools to model narratives and how readers engage with them. Considering the relations between text, reader, and reading situation as constitutive of narrative requires a way of looking at stories keeping in mind that the reader's experiential background, their cognitive-affective states, and the situational context all play a crucial role in the emergence of what we call a narrative. Here I suggest that Large Language Models can be considered artificial systems that simulate a (disembodied) reading process and make it possible to model narrative as a complex system, including the encoded experiential background of the (simulated) reader.

SloLitAA: Slovenian Literary Authorship Attribution via AutoML

Boshko Koloski[1], Andrejka Žejn[2], Senja Pollak[1]

[1] *Jožef Stefan Institute*
[2] *The Research Centre of the Slovenian Academy of Sciences and Arts*

We introduce SloLitAA, the first Slovene benchmark for literary authors attribution, and an AutoML-driven neurosymbolic pipeline that automatically selects effective, interpretable feature-model ensembles for identifying authors. We construct a new dataset containing 5,028 paragraphs from 20 well-known 19th-century Slovene authors, extending the previous literary corpora. We propose three evaluation settings: leave-one-book-out (LOBO), random paragraph split, and train-on-one-book (TOOB), where the goal is to attribute an author based on learning representations from a single book per author. We demonstrate that our AutoML configuration consistently outperforms competitive learners across tasks. Moreover, we demonstrate that by feeding LLMs the most important features based on feature importance, we can generate understandable, human-readable author profiles.

Automatic detection of folkloristic motifs with large language models: the Cinderella tale

Tjaša Arčon[1], Marko Robnik-Šikonja[1], Polona Tratnik[2,3]

[1] *University of Ljubljana, Faculty of Computer and Information Science*
[2] *University of Ljubljana, Faculty of Arts*
[2] *Institute IRRIS*

Artificial intelligence is increasingly applied across many research areas, including digital humanities. We present a novel methodology for automating motif detection in folktales using large language models (LLMs), focusing on motifs defined by the ATU Type index - a classification system used by folklorists to categorize traditional narratives. As a case study, we analysed variants of the Cinderella tale. We evaluated the zero-shot performance of GPT-4.5 Preview on a manually annotated sample of 13 Cinderella folktales or fairy tales and then extended the analysis to 77 variants in English from diverse geographical regions. The model was prompted to detect motifs organized in three sets: (1) 15 basic motifs, (2) 18 extended motifs, and (3) 14 generalized "super-motifs." The resulting motif presence/absence data were transformed into matrices, which served as input for clustering algorithms used in the similarity analysis. Subsequently, we examined how 33 Slovene

Cinderella variants fit within this international framework. To our knowledge, this is the first study using LLMs to detect motifs corresponding to specific ATU types. Our results show that LLMs can effectively detect motifs in a large collection of Cinderella variants and identify complex patterns in folktales. This methodology enables large-scale, multilingual folkloristic analysis and offers new tools for studying traditional narratives computationally.

From Datasets to Data Spaces: What they are and what to expect?

Slavko Žitnik[1],

[1] *University of Ljubljana, Faculty of Computer and Information Science*

We will explore the evolution of the data space concept—from its origins in the European data strategy to emerging reference architectures designed to support sovereign, interoperable, and purpose-specific data ecosystems. We'll examine sectoral implementations across domains such as health, mobility, and culture, and critically assess their technical and societal value. Special focus will be given to the Language Data Space (LDS), a strategic initiative aiming to facilitate large-scale, multilingual language data sharing across Europe. As language data becomes foundational for training robust and transparent AI models, the LDS offers a promising infrastructure for advancing AI research and unlocking new opportunities for the digital humanities.

Teaching Vision-Language Models Semiotics: Toward a Socio-Semiotic Framework for Multimodal AI

Matej Martinc[1], Jan Babnik[2]

[1] *Jozef Stefan Institute*
[2] *Institute IRRIS for Research, Development and Strategies of Society, Culture and Environment*

In our study, we examine the intersection of generative AI, multimodal learning, and socio-semiotic theory, with a particular focus on how Vision-Language Models (VLMs) can be trained to interpret and reason about complex visual meaning. By contrasting machine learning paradigms that treat media as data with social semiotic frameworks that treat modes as meaning-bearing resources, we provide a preliminary assessment of how VLMs engage in inter-semiotic translation and multimodal interpretation. While conventional machine learning approaches often rely on quantitative correlations between visual features and textual outputs, social semiotics (in particular multimodal semiotics) emphasizes the culturally, situationally and contextually embedded, interpretive nature of visual communication. This tension between quantitative and qualitative epistemologies is not merely philosophical, since it manifests in practical challenges, such as hallucination, shallow reasoning, and context-insensitive outputs from contemporary models. Our goal is to explore methodological refinements that bring these paradigms into closer dialogue, allowing for the design of models that are more attuned to the nuances of visual meaning-making. To achieve this goal, we conduct a preliminary qualitative study of how state-of-the-art proprietary VLMs understand semiotics. Based on the findings of the study, we propose a training framework that will enhance semiotic capabilities of VLMs.

AI & Space

Track Chairs

Dino Ienco, National Institute of Agricultural Research
Chiara Maria Cocchiara, European Space Agency
Krištof Oštir, University of Ljubljana

09:00 - 09:50	**Lorenzo Bruzzone** Remote Sensing and the Labeled Data Challenge in the Foundation Model Era
09:50 - 10:10	Pretraining for Dense Prediction Tasks in Earth Observation
10:10 - 10:30	AI-Driven Habitat Change Detection in Alpine Protected Areas: Foundation Models vs. Direct Change Detection
10:30 - 11:00	**Coffee Break**
11:00 - 11:50	**Gilberto Câmara** Machine Learning and AI for Remote Sensing Application: Promises and Challenges
11:50 - 12:10	Towards a Foundation Model for Water Monitoring: Fusing Aligned and Unaligned Multimodal Observations
12:10 - 12:30	Toward Improved Land Surface Forecasts: Foundation Models vs. Task-Specific Approaches
12:30 - 12:50	Deep Learning tools to support deforestation monitoring in the Ivory Coast using SAR and Optical satellite imagery
12:50 - 14:30	**Lunch Break**
14:30 - 14:50	Hierarchical and Explicit Label Modeling with Graph Learning for Multi-Label Remote Sensing Image Classification
14:50 - 15:10	Near Real-Time Flood Mapping from Sentinel Data Using Machine Learning Techniques
15:10 - 15:30	Orthorectification of satellite images with AI-extracted road networks
15:30 - 15:50	Skeleton-Conditioned Diffusion for Interpretable Exoplanet Spectral Retrieval
15:50 - 16:30	**Coffee Break**

Remote Sensing and the Labeled Data Challenge in the Foundation Model Era

Lorenzo Bruzzone[1]

[1] *University of Trento*

Deep neural networks have become the primary approach for extracting semantic information from remote sensing data, supporting a wide range of tasks such as automatic classification, regression, and change detection. Despite the widespread use of supervised learning, a critical limitation in remote sensing applications is the scarcity of reliable labeled data for training deep models. This stands in contrast to the field of computer vision, where large-scale annotated datasets are more readily available. The limited availability of high-quality labeled samples highlights the need for weakly supervised learning strategies capable of effectively integrating multiple sources of imperfect labels into the training process. These weak labels are often associated with various types of uncertainty, which must be carefully modeled to enable robust learning.

This presentation addresses these challenges by offering a critical examination of learning in remote sensing, including the perspective offered by recent advancements in foundation models. It explores the diverse sources of labeled data and the inherent trade-offs between annotation quality and quantity. The main sources of uncertainty in weak labels—such as label noise, temporal mismatch due to land-cover changes, spatial ambiguity, and semantic inconsistency—will be analyzed in detail. Key approaches for robust weakly supervised learning will be discussed, including strategies for both single-date imagery and multi-temporal datasets. Methods such as few-shot learning, scribble-based supervision, and other forms of weak supervision will be explored, along with the potential and limitations of foundation models in reducing dependence on large-scale annotated datasets.

The presentation will illustrate these concepts through applications in climate change monitoring, precision agriculture, and planetary exploration.

Pretraining for Dense Prediction Tasks in Earth Observation

Filip Wolf[1], Blaž Rolih[1], Luka Čehovin Zajc[1]

[1] *University of Ljubljana, Faculty of Computer and Information Science*

Deep learning is becoming an increasingly valuable tool in Earth Observation, with recent advances in self-supervised learning and knowledge transfer leading to the development of foundation models tailored to remote sensing data. However, most existing foundation models are based on masked image modeling, which often underperforms on dense prediction tasks such as segmentation and detection. In this work, we explore alternative pretraining strategies based on contrastive learning, which are better suited for dense tasks. We evaluate our approach on standard benchmarks including GEO-Bench, Sen1Floods11, and SpaceNet. Preliminary results indicate that our method achieves improved quantitative performance and more robust qualitative outputs while requiring less pretraining data.

AI-Driven Habitat Change Detection in Alpine Protected Areas: Foundation Models vs. Direct Change Detection

Harald Kristen[1], Daniel Kulmer[2], Manuela Hirschmugl[1]

[1] *University of Graz*
[2] *Joanneum Research*

Rapid climate change in alpine ecosystems demands frequent habitat monitoring, yet traditional manual mapping approaches are prohibitively expensive for the temporal resolution needed. This study builds on existing long-term alpine habitat data for training purposes. We compare two fundamental change detection (CD) paradigms: (1) post classification CD versus (2) direct CD.

We systematically evaluate modern AI architectures across both paradigms using a unique 20-year alpine habitat dataset from Gesäuse National Park, Austria. For post-classification CD approaches, we compare geospatial foundation models Prithvi-EO-2.0 and Clay v1.0 against traditional U-Net-based CNNs. Direct change detection evaluates the specialized transformer ChangeViT against U-Net baselines.

Using very high resolution multimodal data including RGB, near-infrared (NIR), LiDAR elevation, and terrain attributes across 4,480 documented habitat changes over 15.3 km^2, our analysis reveals key performance differences. Preliminary results show foundation models outperforming CNNs for semantic habitat classification (Clay: 0.60 vs U-Net: 0.43 OA), while direct change detection achieves promising performance for binary classification (IoU 0.48-0.49). Complete paradigm comparison will be enabled by forthcoming 2020 habitat labels.

Machine Learning and AI for Remote Sensing Application: Promises and Challenges

Gilberto Câmara[1,2]

[1] *Fundação Getúlio Vargas*
[1] *EOSTAT, FAO*

Satellite imagery has become invaluable for many applications, ranging from environmental monitoring and urban planning to disaster response and agriculture. Two recent major developments have broadened the use of satellite data: (a) availability of cloud services with large collections of open data, with an emphasis on the Sentinel and Landsat programmes; b) advancement of machine learning (ML) methods to analyse and extract meaningful information from images. The combination of big Earth Observation (EO) data and advanced machine learning for satellite data (SatML in what follows) has substantial potential for supporting evidence-based policy to address global environmental challenges.

There are important shortcomings in most current SatML models. Most techniques do not consider important differences between satellite images and natural ones. Many SatML algorithms, inspired by the U-net paradigm, rely on the distinction between foreground (things) and background (stuff). While this design suits high spatial resolution images with 3-meter or smaller pixels, it is unsuited for mid to low-resolution images (10-meter or larger pixels). Mid and low-resolution images are continuous distributions of radiance values and are better described as fields than as collections of objects. Human-sized, everyday objects depicted in natural photos differ from continuous landscapes captured in satellite images. All pixels matter when working with Sentinel data for land mapping and similar broad-area applications; the foreground/background distinction is not applicable. Arguably, no proper “objects" exist in mid to low-resolution images; image classification identifies compact regions of similar values in multidimensional spaces. While domain scientists may believe they recognise objects in a remotely sensed image, they are actually measuring fields.

A further challenge to SatML models is dealing with satellite image time series. These are calibrated and comparable measures of the same location on Earth at different times. Due to frequent revisits, time series capture both gradual and abrupt changes. Researchers have used time series in applications such as forest disturbance, agricultural production, and land cover mapping. To work with satellite image time series, SatML models must include a temporal component, which is not present in most natural image collections.

This presentation analyses whether the current generation of SatML addresses the ontological challenges of dealing with images of continuous landscapes in different parts of the planet. One case of particular interest is using satellite images to address global environmental change. By covering the same location multiple times, satellites are unique in surveying large areas that are difficult to observe from the ground. Images of tropical and boreal forests, polar regions, and areas at risk of desertification require data and analysis methods that are adequately designed and properly tested. We consider the current status of SatML benchmarks for global change applications. We identify gaps and missing features and propose a way forward so that the community can produce a new generation of SatML datasets and methods adequate for global environmental change.

Towards a Foundation Model for Water Monitoring: Fusing Aligned and Unaligned Multimodal Observations

Blaž Rolih[1], Filip Wolf[1], Alina Machidon[1], Luka Čehovin Zajc[1]

[1] *University of Ljubljana, Faculty of Computer and Information Science*

Surface water monitoring provides valuable support for Earth observation tasks related to ecology, energy, and disaster management. However, real-time monitoring remains challenging due to heterogeneous Earth observation data and a lack of annotations. To address this, we propose a multimodal pretraining framework that fuses aligned drone and satellite imagery while leveraging unaligned samples to improve data efficiency. Preliminary results on alpine river data show improved semantic segmentation performance, with a mean IoU of 88% when both modalities are available. Our approach represents a step towards scalable foundation models for fine-grained water monitoring.

Toward Improved Land Surface Forecasts: Foundation Models vs. Task-Specific Approaches

Marieke Wesselkamp[1], Vitus Benson[1], Markus Zehner[1], Sebastian Hoffmann[1], Claire Robin[1], Melanie Weynants[1], Lazaro Alonso[1], Christian Reimers[1], Nuno Carvailhais[1], Markus Reichstein[1]

[1] *Max-Planck Gesellschaft*

Accurate prediction of land surface dynamics is critical for weather and climate modeling but remains challenging due to the complex, multi-scale nature of Earth system processes and limitations in observational data. This study explores whether geospatial foundation models, trained on diverse Earth observation and reanalysis data, can outperform traditional task-specific AI models for forecasting variables like land surface temperature. By directly comparing these approaches within the EU WeatherGenerator project, we aim to reveal when general-purpose models deliver meaningful advantages — and when specialized solutions remain essential.

Deep Learning tools to support deforestation monitoring in the Ivory Coast using SAR and Optical satellite imagery

Matteo Salis[1], Gabriele Sartor[1], Stefano Pinardi[1], Özgür Saracik[1], Rosa Meo[1]

[1] *University of Turin*

Forest management and deforestation monitoring are pivotal tasks to pursue sustainable development, especially in developing countries where agriculture is the main source of income. In Ivory Coast, for instance, cocoa production is the most remunerative activity, and portions of ancient forests have been deforested in place of new cocoa plantations. To monitor this type of deleterious activity, satellites can be employed to recognize the disappearance of the forest. Here we present a research in which we developed a refined version of the global ALOS Forest-Non-Forest map (FNF) focusing on Ivory Coast, and we compared state-of-the-art Deep Learning (DL) models to predict Forest-Non-Forest segmentation map leveraging Sentinel-1 and Sentinel-2 data. Despite the lack of local forest coverage datasets in Ivory Coast, Sentinel coverage issues, and the challenges posed by persistent cloud cover affecting optical data, we demonstrated the feasibility of using open datasets to build models that classify forest and non-forest pixels in the region.

Hierarchical and Explicit Label Modeling with Graph Learning for Multi-Label Remote Sensing Image Classification

Marjan Stoimchev[1], Boshko Koloski[1], Jurica Levatić[1], Dragi Kocev[1], Sašo Džeroski[1]

[1] *Jožef Stefan Institute*

Hierarchical multi-label classification (HMLC) is essential for modeling complex and structured label dependencies in image analysis domains, particularly remote sensing and medical diagnostics. Despite advances in deep learning and computer vision, existing methods struggle with scenarios where instances belong to multiple paths in hierarchies, failing to fully utilize hierarchical information.

This paper introduces HELM, a novel hierarchical multi-label learning framework addressing this limitation through hierarchical semantic initialization, graph-based structure encoding, and adaptive multimodal fusion. HELM leverages graph convolutional networks to propagate taxonomic relationships while employing an adaptive level-aware training objective that automatically selects appropriate loss functions across hierarchy levels.

To support HMLC research advancement, we construct multi-path label hierarchies for existing datasets spanning remote sensing, medical imaging, and fine-grained visual categorization across nine datasets. Extensive evaluations demonstrate substantial performance improvements, with pronounced gains on complex taxonomic structures across diverse domains. Few-shot experiments reveal that hierarchical structure provides valuable regularization in data-limited scenarios. Results underscore HELM's effectiveness in exploiting hierarchical label relationships, making it particularly valuable for domains requiring precise multi-path classification. The framework's ability to handle complex taxonomic structures across diverse visual domains demonstrates its broad applicability and potential for advancing hierarchical classification tasks in real-world applications.

Near Real-Time Flood Mapping from Sentinel Data Using Machine Learning Techniques

Matija Gerčer[1], Bujar Fetai[1], Tanja Grabrijan[1], Krištof Oštir[1]

[1] *University of Ljubljana (UL FGG)*

In this study, we developed a near-real-time flood mapping service that integrates satellite observations with data from digital elevation models and machine learning techniques. Several publicly available flood-related datasets were evaluated, but none fully met the requirements of our use case in terms of spatial coverage, image quality, or thematic diversity to support robust model development.

To address these limitations, we created a dedicated training dataset based on Copernicus EMS Rapid Mapping products. The dataset includes 40 flood events observed between 2022 and 2025.

A modular workflow was developed to prepare satellite imagery for machine learning applications. It consists of image acquisition, advanced pre-processing methods, flood mask generation, and image tiling. To ensure spatial and temporal consistency across the dataset, additional processing steps—such as rescaling, co-registration, data merging, and masking of non-relevant regions—were implemented.

The preliminary results obtained from the developed models demonstrate promising capabilities in accurately delineating flood extents, with an average IoU of 0.70 on the validation dataset. Although the system is still under active development, these findings highlight the potential for operational deployment.

Orthorectification of satellite images with AI-extracted road networks

Aleš Marsetič[1], Peter Pehani[1], Nina Krašovec[1]

[1] *ZRC SAZU*

High resolution satellite imagery has been available for decades, yet automatic and accurate geometric correction remains a persistent challenge, especially when dealing with imagery which exhibit high radiometric variability. This paper introduces an enhanced version of the geometric processing module within the STORM processing chain, designed to perform fully automated orthorectification of satellite images. The module leverages publicly available ancillary data and deep learning-based road extraction techniques to eliminate the need for manual data collection and preprocessing. Ground Control Points (GCPs) are automatically generated by matching roads extracted from satellite imagery with corresponding vector roads obtained from open-access web databases. The orthorectification pipeline integrates several key components: ancillary data preparation, road detection, GCP extraction, and final orthorectification using a lidar digital terrain model (DTM). Experimental results on WorldView-3 imagery demonstrate that the proposed method achieves sub-pixel Root Mean Square Error (RMSE) accuracy. The integration of deep learning for road detection offers a novel and effective approach for the fully automated orthorectification of satellite data of different types.

Skeleton-Conditioned Diffusion for Interpretable Exoplanet Spectral Retrieval

Ran Liu[1]

[1] *UnibridgeAI*

Recent advances in neural spectral retrieval have greatly accelerated inference but often sacrifice interpretability. Most deep models treat spectra as flat input vectors, obscuring the physical reasoning behind predictions. This paper introduces SAP (Skeleton–Attention–Physics), a structure-aware generative framework that integrates multi-scale spectral skeletons into the conditioning stream of a diffusion model. SAP encodes geometric and topological features, aligns attention with physically salient regions, and enables retrieval via a lightweight decoder. The framework bridges a key gap in explainable retrieval by coherently linking spectral structure, attention dynamics, and physical parameters.

AI & Physics

Track Chairs

Gabrijela Zaharijas, University of Nova Gorica - SMASH supervisor and PI
Borut Paul Kerševan, Jožef Stefan Institute, University of Ljubljana - SMASH supervisor

Tuesday

09:00 - 09:20	A robust neural determination of the source-count distribution of the Fermi-LAT sky at high latitudes
09:20 - 09:40	GPT-like transformer model for silicon tracking detector simulation
09:40 - 10:00	Mixture-of-Expert Autoencoders for Cross-Modality Embedding of Supernovae
10:00 - 10:30	Suitability Analysis of Kolmogorov-Arnold Networks for Real-World Equation Discovery
10:30 - 11:00	**Coffee Break**
11:00 - 11:20	Probing the Parameter Space of Axion-Like Particles Using Simulation-Based Inference
11:20 - 12:00	**Mario Juric** We're Building the Largest Sky Survey in History: Can AI Help us Understand Its Data?
12:00 - 12:20	Towards foundation model for astrophysical source detection: An End-to-End Gamma-Ray Data Analysis Pipeline Using Deep Learning
12:20 - 12:40	Anomalous variability detection in the LSST era
12:40 - 13:00	Nuclear recoil detection with color centers in bulk lithium fluoride
13:00 - 14:30	**Lunch Break**
14:30 - 14:50	Detecting Localized Density Anomalies in Multivariate Data
14:50 - 16:00	**Uroš Seljak** Solving inverse problems in science with AI
16:00 - 16:30	**Coffee Break**
16:30 - 16:50	Modeling Multi-messenger Emission from Galaxy Clusters Using Machine Learning
16:50 - 17:10	Primordial Black Hole formation from a massless scalar field

Wednesday

09:00 - 09:20	Analytic symbolic regression emulators for the impact of baryons on the matter power spectrum
09:20 - 09:40	Simulation-based Inference for LISA: inferring the galactic binary population
09:40 - 10:00	Approximate Bayesian algorithm for tensor-robust PCA using relative entropy
10:00 - 10:30	Alpha-stable noise for establishing secure communications via AI/ML

10:30 - 11:00	**Coffee Break**
11:00 - 11:20	CaloClouds 3: Ultra-Fast Geometry-Independent Highly-Granular Calorimeter Simulation
11:20 - 12:00	**Roberto Trotta** The promise of Simulation-Based Inference for cosmology
12:00 - 12:20	Application of Score-Based Denoising Diffusion Model to Astrophysical Images
12:20 - 12:40	Simulation-based inference has its own Dodelson-Schneider effect (but it knows that it does)
12:40 - 13:00	Diffusion-based models for fast, scalable and accurate baryonic observables predictions beyond power spectra
13:00 - 14:30	**Lunch Break**
14:30 - 14:50	Large-scale reinterpretation of new physics searches at the LHC using ML
14:50 - 16:00	**Gregor Kasieczka** The road to AI-based discoveries
16:00 - 16:30	**Coffee Break**
16:30 - 16:50	Advanced Particle classification in space missions using machine learning techniques
16:50 - 17:10	Advanced Tracking Analysis in Space Experiments with Graph Neural Networks

A robust neural determination of the source-count distribution of the Fermi-LAT sky at high latitudes

Christopher Eckner[1], Noemi Anau Montel[2], Florian List[3], Francesca Calore[4], Christoph Weniger[5]

[1] *University of Nova Gorica*
[2] *Max Planck Institute for Astrophysics*
[3] *University of Vienna*
[4] *LAPTh, CNRS*
[5] *University of Amsterdam*

Over the past 16 years, the Fermi Large Area Telescope (LAT) has significantly advanced our view of the GeV gamma-ray sky, yet several key questions remain - such as the nature of the isotropic diffuse background, the properties of the Galactic pulsar population, and the origin of the GeV excess towards the Galactic Centre. Addressing these challenges requires sophisticated astrophysical modelling and robust statistical methods capable of handling high-dimensional parameter spaces. In this work, we analyse 14 years of high-latitude ($|b| > 30°$) Fermi-LAT data in the 1–10 GeV range using simulation-based inference (SBI) via neural ratio estimation. This approach allows us to detect individual gamma-ray sources and derive a source catalogue with estimated positions and fluxes that are consistent with the bright portion of the Fermi-LAT collaboration's 4FGL catalogue. Additionally, we reconstruct the source-count distribution in both parametric and non-parametric forms, achieving good agreement with previous literature results and detected sources. We also quantitatively validate our gamma-ray emission simulator via an anomaly detection technique, demonstrating that the synthetic data closely reproduces the complexity of the real observations. This study highlights the practical utility of SBI for complex, high-dimensional problems in gamma-ray astronomy. It lays the groundwork for its application to more challenging sky regions or data from next-generation facilities such as the Cherenkov Telescope Array Observatory.

GPT-like transformer model for silicon tracking detector simulation

Tadej Novak[1]

[1] *Jozef Stefan Institute*

Simulating physics processes and detector responses is essential in high energy physics but accounts for significant computing costs. Generative machine learning has been demonstrated to be potentially powerful in accelerating simulations, outperforming traditional fast simulation methods. While efforts have focused primarily on calorimeters initial studies have also been performed on silicon detectors.

This work employs the use of GPT-like transformer architecture in a fully generative way ensuring full correlations between individual hits. Taking parallels from text generation hits are represented as a flat sequence of feature values. The resulting tracking performance, evaluated on the Open Data Detector, is comparable with the full simulation.

Large-scale reinterpretation of new physics searches at the LHC using ML

Judita Mamuzic[1]

[1] *Jozef Stefan Institute*

Searches at the Large Hadron Collider (LHC) for new phenomena beyond the current theory of particle physics, which is known as the Standard Model and describes all known fundamental particles and their interactions, have not produced definitive discoveries. However, several analyses reported mild excesses exceeding 2 standard deviations. These excesses remained largely unexamined in relation to one another. This work applies unsupervised machine learning to extract additional insights from existing LHC searches. The phenomenological Minimal Supersymmetric Model (pMSSM) provides a flexible framework to simulate potential new physics signals. A large-scale reinterpretation of new physics searches is used to identify models consistent with observed excesses. The results of the study aim to highlight directions for future LHC searches.

Suitability Analysis of Kolmogorov-Arnold Networks for Real-World Equation Discovery

Nicolas Kriener[1], Jannis Brugger[1], Mira Mezini[1], Stefan Kramer[2]

[1] *TU Darmstadt*
[2] *Uni Mainz*

Kolmogorov-Arnold Networks (KANs) are a novel deep learning architecture that can be transformed into a closed-form equation after training. We test their applicability as an equation discovery framework on a selection of equations from the Feynman Benchmark and on data measured in a real-world setting with a simulated target variable. Due to their unstable implementation and the ambiguous influence of their hyperparameters, our experiments show that while KANs are a promising approach, they are not yet comparable to state-of-the-art genetic programming frameworks such as PySR.

Probing the Parameter Space of Axion-Like Particles Using Simulation-Based Inference

Pooja Bhattacharjee[1], Gabrijela Zaharijas[1], Christopher Eckner[1]

[1] *University of Nova Gorica*

Axion-like particles (ALPs) appear in various extensions of the Standard Model and can interact with photons, leading to ALP-photon conversions in external magnetic fields. This phenomenon can introduce characteristic energy-dependent "wiggles" in gamma-ray spectra. The Cherenkov Telescope Array Observatory (CTAO) is the next-generation ground-based gamma-ray observatory, designed to provide enhanced sensitivity and energy coverage (20 GeV – 300 TeV) over current Imaging Atmospheric Cherenkov Telescopes (IACTs) and offers an excellent opportunity to study such effects.

In this work, we employ Simulation-Based Inference (SBI) to explore the parameter space of ALPs, targeting the flaring states of blazars, which are among the brightest gamma-ray sources and ideal candidates for probing ALP-induced spectral modulations. Additionally, we investigate whether this inference method can produce accurate ALP exclusion limits comparable to those reported in previous studies that use the classical likelihood-ratio approach. Through this approach, we seek to yield robust constraints on ALP-photon interactions and make substantial advancements in this field.

The promise of Simulation-Based Inference for cosmology

Roberto Trotta[1]

[1] *International School for Advanced Study in Trieste*

On the verge of the step-change in cosmological analysis represented by LSST/Vera Rubin Observatory, traditional inference methods require overhauling in order to deal with the large number of objects and subtle statistical and modelling effects that will otherwise dominate systematics.

I present the case for Neural Ratio Estimation (NRE), a type of Simulation-Based Inference, in the context of supernova type Ia (SNIa) cosmology, showing that NRE matches traditional likelihood-based hierarchical Bayesian modeling on real data; removes systematics offsets due to linearization in large (∼100,000) samples; performs Bayesian model selection at almost no additional computational cost; deals effortlessly with complex selection effects; enables sophisticated calibrations of posterior intervals and confidence regions thanks to its amortized nature. I will demonstrate how the power of SBI can be harnessed by conducting joint inference on supernovae type Ia and their host, obtaining much superior constraining power. Once fully integrated into the data analysis pipeline, NRE has the potential of becoming the tool of choice for SNIa cosmology in the 21st century.

Towards foundation model for astrophysical source detection: An End-to-End Gamma-Ray Data Analysis Pipeline Using Deep Learning

Judit Pérez-Romero[1], Saptashwa Bhattacharyya[1], Sascha Caron[2,3], Dmitry Malyshev[4], Giacomo Principe[5], Roberto Ruiz de Austri Bazán[6], Nicolaas Rodney[2], Zoja Rokavec[7], Danijel Skočaj[7], Domen Tabernik[7], Gabrijela Zaharijas[1]

[1] *University of Nova Gorica*
[2] *Radboud University*
[3] *Nikhef*
[4] *ECAP*
[5] *INFN/INAF*
[6] *IFIC UV-CSIC*
[7] *University of Ljubljana*

The increasing volume of gamma-ray data from space-borne telescopes, like Fermi-LAT, and the upcoming ground-based telescopes, like the Cherenkov Telescope Array Observatory (CTAO), presents us with both opportunities and challenges. Traditional analysis methods based on likelihood analysis are often used for gamma-ray source detection and further characterization tasks. A key challenge to analysing data from these telescopes arises due to background contamination; consistent of interstellar emission for Fermi-LAT and cosmic ray background for CTAO data, which obscures the faint source population.

Here we will present our results from an end-to-end Deep Learning (DL) based pipeline for detection, localization and further characterization tasks of gamma-ray sources. We extend our AutoSourceID (ASID) pipeline, a DL-based pipeline initially tested with Fermi-LAT simulated data and optical data (MeerLICHT), to include results for CTAO simulated data for Galactic Plane Survey (GPS) observation. We will also present a pre-processing step designed with Deep Neural Net for denoising tasks which can potentially decrease the source detection threshold. Training on data from different telescopes to capture a broad representational space of the gamma-ray sky, this end-to-end pipeline - from denoising to detection and characterization - could potentially serve as a foundational model for gamma-ray astronomy by offering a generalizable framework for other surveys.

Anomalous variability detection in the LSST era

Oleksandra Razim[1]

[1] *University of Nova Gorica*

Astronomical datasets have reached the petabyte scale, and the classical methods of scientific discovery that rely on experts looking through hundreds of images, spectra or light curves are not applicable anymore. Nowadays, we routinely use Machine Learning for classifying and characterising astronomical objects, however, novelty detection is falling behind due to the complexity of the task: without ground truth, it is particularly hard to discern between a true anomaly, interesting from the physical point of view, and a bogus one. The task is better solved for image-based analysis, but the variability analysis using light curves is a less explored area.

Space-based telescopes, such as TESS and Kepler, discovered that about 60% of the stars can be considered as variables when measured with millimag photometric precision. With the arrival of the Legacy Survey of Space and Time (LSST) which will observe billions of stars with photometric accuracy of an order of 10 mmag or better, and subsequent start of operations of the Roman telescope in the next few years, we face the challenge of finding novelties in a huge, highly heterogeneous and very contaminated dataset, otherwise we risk missing an insignt into stellar evolution or even a hint to the new physics. In this talk, we present MapLC, a project dedicated to the development of a semi-automatic model-agnostic anomaly discovery pipeline for variable astronomical sources, such as rare types of stars or previously unseen subclasses of Tidal Disruption Events. We discuss which areas of the parameter space hold the biggest promise of discovery with the arrival of the new generation of telescopes, and how we can utilise these discoveries to shed light on the obscure aspects of stellar astrophysics and the evolution of our Galaxy.

Nuclear recoil detection with color centers in bulk lithium fluoride

Patrick Stengel[1]

[1] *Jožef Stefan Institute*

We present initial results on nuclear recoil detection based on the fluorescence of color centers created by nuclear recoils in lithium fluoride. We use gamma rays, fast and thermal neutrons, and study the difference in responses they induce, showing that this type of detector is rather insensitive to gamma rays. We use light-sheet fluorescence microscopy to image nuclear recoil tracks from fast and thermal neutron interactions deep inside a cubic-centimeter sized crystal and demonstrate automated feature extraction in three dimensions using machine learning tools. The number, size, and topology of the events agree with expectations based on simulations with TRIM. These results constitute the first step towards 10-1000g scale detectors with single-event sensitivity for applications such as the detection of dark matter particles, reactor neutrinos, and neutrons.

Detecting Localized Density Anomalies in Multivariate Data

Sebastian Springer[1], Andre Scaffidi[1], Max Autenrieth[2], Gabriella Contardo[3], Alessandro Laio[1], Roberto Trotta[1], Heikki Haario[4]

[1] *SISSA*
[2] *Cambridge University*
[3] *University of Nova Gorica*
[4] *LUT*

We introduce a methological pipeline for detecting local density anomalies between two multivariate sets of points. Unlike traditional outlier detection, which targets individually unusual points in a given dataset, we aim to identify whether–and where–two datasets differ

in local point density, even when those points appear typical. This task is crucial in various fields: when comparing observed data with theoretical expectations (as in particle physics, to detect particles that are not present in the "standard model"), when examining different underlying populations (e.g. the properties of stars in the Milky Way disk versus the halo), or when analyzing data from different periods (such as climate patterns now versus 20 years ago).

To address this problem, we propose a distribution-free approach that avoids explicitly modelling either dataset. First, we define an anomaly score that is computed for each datapoint using its nearest-neighbors' memberships (i.e. which dataset a point is from). A large anomaly score is obtained if the neighbors' memberships deviate from a Binomial distribution, indicating an over- or under-density compared to the counter-part dataset. From these scores, we can thus identify the sets of points lying in anomalous state-space. We also derive "equalized" datasets, i.e. two subsets of points from each original dataset that are now indistinguishable under our statistic, and present an injection scheme that allows the estimation of the anomalous (signal) vs non-anomalous ("background") component ratio. We demonstrate the effectiveness of our pipeline on controlled synthetic datasets and real-world challenges, such as the anomaly detection benchmark task in simulated Large Hadron Collider data, and climate temperature data.

Solving inverse problems in science with AI

Uroš Seljak[1]

[1] *University of California, Berkeley*

Scientific inverse problems in science are some of the most difficult problems characterized by high dimensionality (often 2d or 3d images), and expensive forward models. In recent years there has been a lot of development of AI powered solutions to inverse problems. Two of the most common approaches are explicit and implicit likelihood analyses. In explicit likelihood analysis one performs optimization or Monte Carlo Markov Chain (MCMC) sampling, and there has been significant progress in methods development, including novel gradient based MCMC methods such as MicroCanonical Langevin Monte Carlo. Implicit likelihood analysis, also called Simulation Based Inference, uses forward model simulations and Neural Networks to learn the likelihoods, and this approach has been shown to be able to solve problems that were previously not feasible. I will present examples of both approaches to data analysis in cosmology.

Modeling Multi-messenger Emission from Galaxy Clusters Using Machine Learning

Saqib Hussain[1]

[1] *University of Nova Gorica*

We calculate the high-energy gamma-ray and neutrino emissions from galaxy clusters like Perseus that host active galactic nuclei (AGNs). Our primary objective is to distinguish the emission from the central source, such as NGC1275, from the diffuse emission originating in the outskirts of the Perseus cluster. We employ three-dimensional cosmological magnetohydrodynamical simulations of structure formation to model the turbulent intracluster medium (ICM). We propagate CRs in the ICM and intergalactic medium using multi-dimensional Monte Carlo simulations, considering all relevant photohadronic, photonuclear, and hadronuclear interactions. By comparing our results with the existing upper limits from IceCube for galaxy clusters and the sensitivity of CTA, we predict that these observatories could potentially establish a new class of astrophysical sources capable of emitting high-energy multi-messenger signals. Based on our findings for nearby clusters

such as Perseus and Coma, we are training a machine learning model to forecast the high-energy emission spectra of galaxy clusters across a range of higher redshifts.

Primordial Black Hole formation from a massless scalar field

Ilia Musco[1]

[1] *University of Nova Gorica*

We have investigated primordial black holes formation within a model of the early Universe dominated by a massless scalar field, developing a numerical spherically symmetric code dedicated to this problem. Imposing initial conditions on super horizon scales using the gradient expansion approach, we show that a massless scalar filed is equivalent to a perfect fluid, where the pressure is equal to the total energy density, only in the regime of cosmological expansion, while during the gravitational collapse space-like gradients arises and the comoving slicing is failing. For this reason, moving to a constant mean curvature slicing, we follow entirely the numerical evolution of cosmological perturbations, also during the gravitational collapse, computing the threshold for primordial black hole formation, and the corresponding mass distribution described by critical collapse. Together with previous simulations of primordial black formation in a radiation dominated Universe, this work will be the basis to expand the study of primordial black holes using machine learning techniques.

Analytic symbolic regression emulators for the impact of baryons on the matter power spectrum

Lukas Kammerer[1], Deaglan J. Bartlett[2], Gabriel Kronberger[1], Harry Desmond[3], Pedro G. Ferreira[2]

[1] *University of Applied Sciences Upper Austria*
[2] *Astrophysics, University of Oxford*
[3] *University of Portsmouth*

Baryonic physics has a considerable impact on the distribution of matter in our Universe on scales probed by current and future cosmological surveys. We seek simple symbolic parametrisations for the impact of baryonic physics on the matter power spectrum for a range of physically motivated models, as a function of wavenumber, redshift, cosmology, and parameters controlling the baryonic feedback.

We use genetic programming and symbolic regression to construct analytic approximations for the ratio of the matter power spectrum in the presence of baryons to that without such effects. We obtain separate functions of each of four distinct sub-grid prescriptions of baryonic physics from the CAMELS suite of hydrodynamical simulations (Astrid, IllustrisTNG, SIMBA and Swift-EAGLE) as well as for a baryonification algorithm. We also provide functions which describe the uncertainty on these predictions, due to both the stochastic nature of baryonic physics and the errors on our fits.

The error on our approximations to the hydrodynamical simulations is comparable to the sample variance estimated through varying initial conditions, and our baryonification expression has a root mean squared error of better than one percent, although this increases on small scales.

These errors are comparable to those of previous numerical emulators for these models. Our expressions are enforced to have the physically correct behaviour on large scales and at high redshift. Due to their analytic form, we are able to directly interpret the impact of varying cosmology and feedback parameters, and we can identify parameters which have little to no effect.

Simulation-based Inference for LISA: inferring the galactic binary population

Rahul Srinivasan[1]

[1] *SISSA*

The upcoming space-based gravitational wave detector, LISA, holds great potential for detecting a wide range of sources, the most numerous of which are galactic binaries (GBs). However, inferring their population poses significant challenges for traditional Bayesian methods due to the vast number of overlapping sources (30×10^6) within LISA's extensive data set (6×10^6 frequency bins). In this talk, I will present a simulation-based inference framework designed to address these challenges. Our approach enables hierarchical inference of GB population characteristics. We leverage GPU-accelerated forward simulators to efficiently generate synthetic LISA data, exploring inference efficacy under a variety of astrophysical models. To handle the high dimensionality of the data, we develop a robust summary statistic that preserves essential population information while significantly compressing the input data. Our results demonstrate the potential for fast, scalable, and accurate inference from LISA data, paving the way for deeper insights into the properties of GBs and the structure of the Milky Way.

Advanced Tracking Analysis in Space Experiments with Graph Neural Networks

Federica Cuna[1], Maria Bossa[1], Fabio Gargano[1], Mario Nicola Mazziotta[1]

[1] *INFN*

The integration of advanced artificial intelligence (AI) techniques into astroparticle physics represents a transformative shift in how data analysis and experimental design are approached. As space-based missions become increasingly complex, leveraging AI has become essential for optimizing performance and achieving robust scientific results.

In this work, we present the development of innovative AI-driven algorithms for particle tracking, focusing on the use of Graph Neural Networks (GNNs)—a class of geometric deep learning models particularly well-suited to the graph-like structure of tracking systems. In this context, detector hits are represented as nodes, while their geometric relationships are encoded as links. GNNs provide a powerful framework for solving tasks such as node classification, edge prediction, and full graph inference, making them ideal for handling the challenges of particle reconstruction in space experiments.

One of the major challenges in space-based tracking systems is the high-noise environment, especially due to backscattered tracks from calorimeters, which obscure the identification of the primary particle trajectory. To address this, we propose a GNN-based node classification algorithm capable of distinguishing signal hits from backscattering hits, thereby enabling the accurate reconstruction of particle tracks. Our approach has been applied on simulated data from a real astroparticle experiment, ensuring that the algorithm is tailored to realistic detector conditions and configurations. The model effectively identifies the hits associated with the primary particle and enables the reconstruction of key track parameters with high fidelity.

By tackling the intrinsic complexity of the space detector environment using cutting-edge AI methodologies, this work contributes to improving the precision and accuracy of data analysis in astroparticle physics, ultimately supporting more insightful scientific discoveries.

Advanced Particle classification in space missions using machine learning techniques

Maria Bossa[1], Federica Cuna[1], Fabio Gargano[1]

[1] *INFN*

The integration of advanced artificial intelligence (AI) techniques into astroparticle experiments marks a transformative step in both data analysis and experimental design. As space missions grow increasingly complex, the adoption of AI technologies becomes critical for optimizing performance and achieving robust scientific outcomes. In this context, we explore innovative AI-driven approaches tailored for space-based calorimetric detectors.

We propose a fully custom-designed Transformer-based model for particle identification in space calorimeters. A key challenge in these experiments is the distinction between particle types, such as electrons and protons, based on energy deposition patterns. By capturing long-range dependencies across thousands of input channels, Transformers offer a powerful framework for robust classification. Our approach aims to enhance both the accuracy and reliability of particle identification, with the potential to extend classification capabilities across a wide energy spectrum, from 1 GeV to 1 TeV.

In parallel, we investigate complementary machine learning techniques trained on high-level reconstructed variables for the classification of electromagnetic and hadronic showers. These traditional ML models, based on physically motivated features, provide an interpretable and computationally efficient alternative that complements the deep-learning strategy.

By combining these approaches, our work highlights the synergy between deep and traditional machine learning methods in advancing calorimetric reconstruction and classification, thereby demonstrating the significant impact of state-of-the-art AI methodologies on the design and performance of future space-based astroparticle experiments.

CaloClouds 3: Ultra-Fast Geometry-Independent Highly-Granular Calorimeter Simulation

Henry Day-Hall[1], Thorsten Henrik Buss[1], Frank Gaede[1], Gregor Kasieczka[1], Katja Krüger[1], Anatolii Korol[1], Peter Mckeown[2], Lorenzo Valente[3]

[1] *DESY*
[2] *CERN*
[3] *University of Hamburg*

This contribution presents the final iteration of the CaloClouds series. Simulation of photon showers in the granularities expected in a future Higgs factory is computationally challenging. A viable simulation must capture the find details exposed by such a detector, while also being fast enough to keep pace with the expected rate of observations. The Caloclouds model utilises point cloud diffusion and normalising flows to replicate MCMC simulation with exceptional actuary.

The improvements made are exhaustive, and we have confidence that this represents the best possible performance for the current architecture. With significant work, we now have a global implementation, for use across the whole calorimeter. This allows us to finalise the series by demonstrating faithful, higher level reconstructed quantities against Geant4 simulation.

First we will make a lightning overview of the models objectives and constraints. To describe the upgrades for the latest version, we detail the studies on the flow model and the optimisations made, and then summarise the steps taken to generalise CaloClouds 3 for using in the whole detector. Considering some of the underlying principles of model design, we look at the significance of the data format choice on model outcomes. Finally, we present the results of reconstructions performed on CaloClouds 3 output against the results from Geant4 simulation, thus demonstrating that this model provides reliable physics reproductions.

We're Building the Largest Sky Survey in History: Can AI Help us Understand Its Data?

Mario Juric[1]

[1] *University of Washington*

In this talk I'll describe how modern machine learning and AI connect with the challenges and opportunities of the Rubin Observatory's Legacy Survey of Space and Time (LSST). LSST — the largest optical sky survey in history, expected to begin by the end of this year — will produce data at unprecedented rates and volumes. I outline how AI can support it every stage of the data-to-science pipeline: from image calibration and difference imaging, to real-time alert characterization, cross-matching, anomaly detection, and large-scale discovery. I'll highlight applications in time-domain astrophysics and Solar System science, and emphasize the emerging roles of community brokers, foundation models, and domain-informed architectures in transforming raw data streams into reliable classifications and actionable follow-up. My goal is to show how collaboration between astronomers and ML researchers can help turn LSST's data deluge into genuine scientific insight.

Application of Score-Based Denoising Diffusion Model to Astrophysical Images

Saptashwa Bhattacharyya[1], Gabrijela Zaharijas[1], Judit Perez Romero[1]

[1] *University of Nova Gorica*

Increasing number of ground-based and space-based telescopes at various wavelengths are producing enormous amounts of data to help us understand our universe better. However, the data analysis pipeline of these telescopes encounter a limitation when the signal-to-noise-ratio is low. At various wavelengths the backgrounds can vary from instrumental to various astrophysical processes that hinders source detection and further characterization processes.

Here, we introduce score-based denoising diffusion model with attention U-Net as backbone to denoise astrophysical images. Using astrophysical data at two different wavelengths obtained from ground-based MeerLICHT and space-based Fermi-LAT telescope, we show that our background removal technique enhances the source detection and subsequent characterization pipeline. We show that irrespective of the density of the field or the background noise level, our method performs better than standard background removal techniques via SourceExtractor. This method will be introduced as a pre-processing step in our complete ML-based AutoSourceID pipeline that detects, localizes and further characterizes the astrophysical sources.

Simulation-based inference has its own Dodelson-Schneider effect (but it knows that it does)

Jed Homer[1], Oliver Friedrich[1], Daniel Gruen[1]

[1] *LMU*

Making inferences about physical properties of the Universe requires knowledge of the data likelihood. A Gaussian distribution is commonly assumed for the uncertainties with a covariance matrix estimated from a set of simulations. The noise in such covariance estimates causes two problems: it distorts the width of the parameter contours, and it adds scatter to the location of those contours which is not captured by the widths themselves. For non-Gaussian likelihoods, an approximation may be derived via Simulation-Based Inference (SBI). It is often implicitly assumed that parameter constraints from SBI analyses,

which do not use covariance matrices, are not affected by the same problems as parameter estimation with a covariance matrix estimated from simulations. We investigate whether SBI suffers from effects similar to those of covariance estimation in Gaussian likelihoods. We use Neural Posterior and Likelihood Estimation with continuous and masked autoregressive normalizing flows for density estimation. We fit our approximate posterior models to simulations drawn from a Gaussian linear model, so that the SBI result can be compared to the true posterior. We test linear and neural network based compression, demonstrating that neither methods circumvent the issues of covariance estimation. SBI suffers an inflation of posterior variance that is equal or greater than the analytical result in covariance estimation for Gaussian likelihoods for the same number of simulations. The assumption that SBI requires a smaller number of simulations than covariance estimation for a Gaussian likelihood analysis is inaccurate. The limitations of traditional likelihood analysis with simulation-based covariance remain for SBI with a finite simulation budget. Despite these issues, we show that SBI correctly draws the true posterior contour given enough simulations.

Diffusion-based models for fast, scalable and accurate baryonic observables predictions beyond power spectra

Satvik Mishra[1], Roberto Trotta[1], Matteo Viel[1]

[1] *SISSA*

Modeling the distribution of neutral hydrogen is essential for understanding the physics of structure formation and the nature of dark matter, but accurate numerical simulations are computationally expensive. We describe novel diffusion-based strategies aimed at removing this computational bottleneck.

First, we describe our Variational Diffusion Model, LODI, trained with the CAMELS simulation suite to generate accurate 21 cm intensity maps from computationally inexpensive dark-matter-only N-body simulations. We demonstrate that LODI delivers highly accurate 21 cm power spectra up to scales as small as $k = 10$ h Mpc^{-1}.

Second, we describe a conditional in-painting technique that is used to synthesize multi-scale baryonic information across multiple observational channels. Shown only the dark-matter and gas temperature in a single 25 Mpc h^{-1} patch, the model produces consistent maps across the whole TNG-300 volume (205 Mpc h^{-1}), preserving the correct one- and two-point statistics without retraining. Together, these capabilities showcase a fast, end-to-end trainable emulator for survey-scale 21 cm intensity maps for arbitrarily large volumes, with the aim of enabling massive forward-modeling capabilites and robust inference of cosmological and dark-matter parameters for experiments such as HERA and the SKA.

Mixture-of-Expert Autoencoders for Cross-Modality Embedding of Supernovae

Alexander Gagliano[1], Yunyi Shen[2]

[1] *The NSF Institute for Artificial intelligence and Fundamental Interactions*
[2] *MIT*

Time-domain astrophysics relies on heterogeneous, multi-modal data. Specialized models are constructed to extract information from each modality, but this ignores a wealth of cross-modality information. In our talk, we will present a multi-modal, mixture-of-expert variational autoencoder (MMVAE) we have developed to learn a joint embedding for supernova light curves and spectra. Our method is inspired by the perceiver architecture and natively accommodates variable-length inputs of irregularly-sampled observations. We

train our model on radiative transfer simulations and validate its performance on cross-modality reconstruction of supernova spectra and physical parameters from the simulation. Our model achieves superior performance to nearest-neighbor searches in a contrastively-trained latent space, showing its promise for constructing informative latent representations of multi-modal astronomical datasets.

The road to AI-based discoveries

Gregor Kasieczka[1]

[1] *University of Hamburg*

Modern machine learning and artificial intelligence fundamentally changing how we analyze huge volumes of data in particle physics and adjacent scientific disciplines. These breakthroughs promise new insights into major scientific questions such as the nature of dark matter or the existence of physical phenomena beyond the standard model.

This keynote will provide an overview of recent, exciting developments with a focus on model agnostic discovery strategies (including first experimental results!), boosting model performance by incorporating physics knowledge, ultra-fast surrogate simulations, as well as foundation models that simultaneously solve multiple tasks across multiple datasets.

Alpha-stable noise for establishing secure communications via AI/ML

Areeb Ahmed [1]

[1] *University of Ljubljana*

Investigations to increase the covertness by strengthening the security of physical layer began in 1950s. Spread spectrum communication was the first step to provide covertness in the communication systems based on the physical layer. Later on, chaotic communication laid the basics of chaotic carrier-based communication systems. The inspiration of chaotic communication arose from the major advantages provided by chaotic signals which have inherent resistance to eavesdropping. However, afterwards, different techniques were also presented to infiltrate chaotic communication systems as well. Recently, a new class of communication system based on random carrier or random noise, i.e. random communication systems, has emerged. Random communication systems provide more covertness, as it is based on stochastic processes which have infinite variance, hence, making it impossible for eavesdropping. Moreover, incorporating machine learning algorithms within the random communication system is an interesting opportunity. In this paper, the authors have provided brief insight into these communication systems and have tried to compare them on common grounds. Additionally, common problems and future challenges related to the physical layer-based communication systems have also been highlighted.

Approximate Bayesian algorithm for tensor-robust PCA using relative entropy

Andrej Srakar[1]

[1] *University of Ljubljana*

Approximate Bayesian (ABC) algorithms are often used to address high-dimensional problems in many areas of physics, in particular cosmology and particle physics. I develop an ABC algorithm for a type of tensor completion where it has still been studied seldom. Matrix completion and, as their extension, tensor completion methods are gaining interest in statistics and machine learning. They allow to study un- and semistructured datasets of some contemporary endeavours such as citizen science initiatives (Robin et al., 2020). Recently proposed Tensor Robust Principal Component Analysis (TRPCA, Lu et al., 2019) aims to exactly recover the low-rank and sparse components from their sum. We extend an own Bayesian approximate inference algorithm for TRPCA (Srakar, 2022), based on regression adjustment methods (Blum, 2010) and compare it to earlier studies using variational Bayes inference. As the estimation is set in a high-dimensional context this leads to known bottlenecks which we solve using our novel proposal to use functional Bregman divergence between posterior distributions as a measure of the posterior surrogate loss, which builds, develops and extends ABC-CDE proposal for complex high-dimensional data (Izbicki et al., 2019; Frigyik et al., 2008). Namely, we use relative entropy as divergence measure combined with more general regression adjustment perspective. We develop proofs on posterior consistency, limiting posterior shape, and asymptotic normality of the posterior mean which is very interesting from the developing area of asymptotics for Bayesian machine learning. We study performance of the approach also in a Monte Carlo simulation study. Our novel approach allows many extensions to modelling phenomena in different areas of physics and we discuss them in conclusion, as well as extension of the approach in the line of future development of matrix completion approaches in high-dimensional statistics and data analysis.

AI & Life Sciences

Track Chairs

Aleksander Sadikov, University of Ljubljana - SMASH supervisor
Jovan Tanevski, Heidelberg University Hospital
Gregor Štiglic, University of Maribor

Dynamics from static data: the example single-cell RNA-seq

Guido Sanguinetti[1]

[1] *SISSA*

Single-cell transcriptomic is a revolutionary way to measure the cell's molecular state, but its destructive nature limits its potential to inform models of cellular dynamics. Recently, La Manno and collaborators introduced the concept of RNA velocity which dramatically changed this assessment, showing that it is indeed possible at least in principle to determine some dynamical information from static data. In this talk, I will introduce the concept of RNA velocity, and show how we expanded it using concepts from machine learning and physics-informed neural networks.

Explainable AI for Clinical Decision Support in Multimorbidity: Toward Transparent and Personalized Care

Martin Michalowski[1]

[1] *University of Minnesota*

Managing multimorbidity presents one of the most pressing challenges in modern healthcare, requiring nuanced clinical decisions that balance complex comorbid conditions, medications, and patient preferences. In this talk, I will explore how explainable AI (XAI) can enhance Clinical Decision Support Systems (CDSSes) to deliver transparent, trustworthy, and personalized guidance to clinicians working with patients who face multiple chronic conditions. Emphasis will be placed on integrating diverse biomedical data into AI models that not only predict outcomes but also offer rationales for their recommendations. Through clinical use cases, I will demonstrate how explainability is crucial for clinical adoption, ethical alignment, and improved patient outcomes. The talk will also highlight design considerations for human-AI collaboration and discuss future directions for building adaptive, interpretable CDSS platforms that support individualized care across varied healthcare settings.

Stimulus-Response Multimodal Emotion Recognition Using EEG and Audio-Visual Data with Cross-Modal Attention

Hyoung-Gook Kim[1], Jörn Fischer[2], Jin-Young Kim[3]

[1] *Kwangwoon University*
[2] *Mannheim University of Applied Sciences*
[3] *Chonnam National University*

This paper presents an enhanced multi-modal emotion recognition framework that combines emotion-eliciting audiovisual stimuli with corresponding EEG responses. Unlike earlier methods that focused solely on either external stimuli or internal responses, our approach jointly analyzes video content and brain activity to provide a more comprehensive and interpretable understanding of emotions. The framework makes three key contributions. First, it models the causal link between audiovisual stimuli and neurophysiological responses by synchronizing audiovisual input with EEG signals, enhancing both accuracy and interpretability. Second, it employs modality-specific encoders: a 3D Shearlet transform and a Vision Transformer for visual features, a VGG19-based CNN for mel-spectrograms of audio, and an EEG Conformer that combines CNN and Transformer modules to capture spatial-temporal brain patterns. Third, a cross-modal attention fusion mechanism

selectively emphasizes complementary features and mitigates redundancy to enable robust emotion classification. The training procedure consists of three stages: pre-training each encoder independently; fine-tuning the cross-modal attention module jointly with the emotion classifier; and testing the fully trained model on the evaluation set. Experimental results on DEAP and SEED datasets demonstrate that the proposed method significantly outperforms single-modality models and conventional fusion approaches, thereby validating the effectiveness of the proposed cross-modal attention strategy for emotion recognition. Acknowledgments: This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. NRF-2023R1A2C1006756).

Exploration of EEG-Based Neural Population Models Using A Probabilistic Grammar

Nina Omejc[1], Sabin Roman[1], Ljupčo Todorovski[2], Sašo Džeroski[1]

[1] *Jožef Stefan Institute*
[2] *Faculty of Mathematics and Physics, University of Ljubljana*

Neural populations often act in synchrony to generate behavior, yet a formal description of these dynamics remains challenging. Existing population models capture specific aspects of neural activity but offer limited guidance on model selection and applicability. To address this, we leverage equation-discovery methods from machine learning to formalize a grammar capable of generating diverse neural population models. The grammar encodes essential structural components—populations within nodes and their connectivity—allowing sampling of single-node models or embedding them into large-scale brain networks. Derived from 16 models in the literature, the grammar expands the design space to an estimated 10^{30} possible variants. A network of nodes, defined by coupled stochastic and delay differential equations, can be simulated and fitted to electroencephalographic (EEG) data via optimization-based parameter estimation. By embedding model construction within the equation-discovery paradigm, our approach facilitates large-scale, data-driven model selection and offers a step toward closer integration of AI with theoretical neuroscience.

Foundation Models for EEG: Pre-training and Fine-tuning

Sebastjan Kramar[1], Gašper Slapničar[1], Lukas Lechner[2]

[1] *Jožef Stefan Institute*
[2] *Austrian Institute of Technology Gmbh*

Electroencephalography (EEG) is difficult to model with deep neural networks due to noise, inter-subject variability and scarce labels. We present a compact pipeline toward an EEG foundation model (FM), comprising the following key components: (i) reproducible preprocessing for TUH and NMT Scalp datasets, (ii) contrastive pre-training with InfoNoise-Contrastive Estimation (InfoNCE) of three UNet variants – base-line, Convolutional Block Attention Module (CBAM) and self-attention– and (iii) task-specific fine-tuning with focal loss. A 12-layer UNet (6 encoder + 6 decoder convolutions with skip connections) trained with Fully Sharded Data Parallel (FSDP) on four GPUs achieved AU C = 0.65 and F1 = 0.49 ± 0.02 on the NMT Scalp dataset, classifying normal and abnormal recordings. CBAM reached AU C = 0.61 and F1 = 0.46 ± 0.02 on the same task. Lightweight spatial attention helped improve EEG feature learning, however absolute performance remained below clinical utility.

Enhancing Parkinson's Disease Diagnosis with Hypomimia and Speech Analysis

Justyna Skibinska[1]

[1] *University of Ljubljana, Faculty of Computer and Information Science*

As society ages, there is a growing need for simple and accessible methods to diagnose and monitor neurodegenerative disorders such as Parkinson's disease (PD). Such tools would allow clinicians to tailor treatments more effectively and improve patients' quality of life. This study investigates the potential of combining computerised assessments of hypomimia and hypokinetic dysarthria to develop a novel diagnostic approach.

The research involved 73 PD patients and 46 healthy controls (age and gender matched), who completed various speech tasks. Acoustic features related to phonation, articulation, and prosody were extracted from voice recordings, while facial movement was analysed using landmark tracking. Both datasets were then processed using the XGBoost algorithm. Acoustic analysis achieved a balanced accuracy 77% in PD detection, while the facial analysis-based support system methodology reached 81%. Combining both modalities improved performance to 83% balanced accuracy (88% sensitivity, 78% specificity). Among speech tasks, tongue twisters were the most clinically informative for the multimodal system.

The trained model based on the tongue twister reached 74 % classification accuracy. The study also provides clinical interpretations of the models. This computer-assisted approach could serve as a supplementary tool for neurologists in the diagnosis of PD.

Machine Learning Models for Predicting Suicidal Ideation, Depression, Anxiety, and Stress in the General Population

Teodora Matić[1], Aleksander Sadikov[1], Peter Pregelj[2], Polona Rus Prelog[2]

[1] *University of Ljubljana, Faculty of Computer and Information Science*
[2] *University Psychiatric Clinic Ljubljana, Centre for Clinical Psychiatry, Ljubljana, Slovenia*

Suicide is a leading cause of death globally, with over 700,000 registered deaths and an estimated 25 million attempts annually. In Slovenia, suicide rates exceed the EU average, with a mortality rate of 17.08 per 100,000 in 2018. Despite extensive research, accurately predicting individuals at risk of suicide remains a challenge, with current tools showing prediction accuracy no better than chance. Depression, anxiety, and stress significantly contribute to mental health burdens, often going underdiagnosed due to overlapping symptoms and subjective screening tools. Machine learning (ML) offers promising advancements in suicide prevention by enabling data-driven predictive models. Building on our prior work predicting suicidal ideation (SI) using ML, this study incorporates depression, anxiety, and stress prediction into the same framework. Using logistic regression models trained on data from a Slovenian population-based study, we achieved AUROC values of 0.84, 0.80, and 0.82 for predicting depression, anxiety, and stress, respectively, alongside 0.82 for SI. Key predictors included behavioral disengagement, self-blame, and denial, while protective factors such as satisfaction with relationships were strongly associated with reduced risk. This model, requiring less than 5 minutes to administer, allows for early identification of individuals experiencing mental distress, facilitating timely intervention. Given the substantial burden of mental health challenges and their association with suicidality, integrating ML-based tools into public health strategies could significantly enhance early detection and intervention efforts, ultimately reducing preventable deaths.

Prevalent Frequency of Emotional and Physical Symptoms in Social Anxiety using Zero Shot Classification

Muhammad Rizwan[1]

[1] *University of Ljubljana*

Social anxiety represents a prevalent challenge in modern society, affecting individuals across personal and professional spheres. Left unaddressed, this condition can yield substantial negative consequences, impacting social interactions and performance. Further understanding its diverse physical and emotional symptoms becomes pivotal for comprehensive diagnosis and tailored therapeutic interventions. This study analyze prevalence and frequency of social anxiety symptoms taken from Mayo Clinic, exploring diverse human experiences from utilizing a large Reddit dataset dedicated to this issue. Leveraging these platforms, the research aims to extract insights and examine a spectrum of physical and emotional symptoms linked to social anxiety disorder. Upholding ethical considerations, the study maintains strict user anonymity within the dataset. By employing a novel approach, the research utilizes BART-based multi-label zero-shot classification to identify and measure symptom prevalence and significance in the form of probability score for each symptom under consideration. Results uncover distinctive patterns: "Trembling" emerges as a prevalent physical symptom, while emotional symptoms like "Fear of being judged negatively" exhibit high frequencies. These findings offer insights into the multifaceted nature of social anxiety, aiding clinical practices and interventions tailored to its diverse expressions.

Insomnia in Patients with Cognitive Decline

Tilen Gabršček[1], Nika Jerman[2], Aleksander Sadikov[1], Polona Rus Prelog[2]

[1] *University of Ljubljana, Faculty of Computer and Information Science*
[1] *University Psychiatric Clinic Ljubljana, Centre for Clinical Psychiatry, Ljubljana, Slovenia*

Sleep disturbances are highly prevalent among elderly individuals and have been associated with cognitive decline and neurodegenerative conditions. We collected data from 60 hospitalized patients over the age of 65 to explore the relationship between cognitive decline and sleep problems. Two potential outcome measures were considered: subjective assessment using the Insomnia Severity Index (ISI) and objective evaluation based on hospital staff observations. Since these measures did not correlate strongly, we analyzed them separately. Patients were categorized as either “sleep well” or “don’t sleep well” according to each method. Overall, 45.8% were classified as “don’t sleep well” based on the ISI and 47.5% based on staff observations. One patient was excluded due to missing data in one of the outcome measures. We first performed univariate regression on all variables and selected those with potential relevance ($p < 0.25$) for inclusion in stepwise regression analysis. Stepwise regression revealed that subjective variables, such as VAS pain score and a history of sleep disturbances, were more strongly associated with self-reported insomnia (ISI). In contrast, objective variables, such as age and the presence of severe cardiovascular disease, were more strongly associated with staff-based assessments of sleep quality. Cognitive decline also showed a correlation with sleep problems, although its effect was overshadowed by other factors in multivariate analysis. While these findings are consistent with previous research, the small sample size warrants further investigation.

Metric Sensitive Loss for Cardiac Segmentation

Muniba Ashfaq[1], Zoran Bosnić[1]

[1] *University of Ljubljana, Faculty of Computer and Information Science*

Accurate segmentation of the heart substructures is crucial to identifying heart diseases. The main substructures of the heart include the left ventricle (LV), myocardium (MYO), and right ventricle (RV). Deep learning techniques have been effective in segmenting these substructures. The loss function is an important component during the training of deep learning models that measures the difference between the ground truth and the predicted outputs. In the segmentation problem, where segmenting objects is considered a pixel-wise classification problem, cross entropy (CE) is the most common loss function. The performance metric considered for evaluating segmentation problem performance is usually region-based Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) or Jaccard index. The main limitation of using cross entropy as a loss function lies in the situation where it starts underweighting the crucial foreground objects with small regions as compared to the background or other classes. The loss functions based on the segmentation performance metrics are better as compared to cross entropy and weighted cross entropy. Moreover, weighted loss functions, i.e., Tversky loss, focal loss, and combo loss, are also used to deal with class imbalance. The Automated Cardiac Diagnosis Challenge (ACDC) dataset contains 100 patients as the training data and 50 patients in the test data. The deep learning model for segmenting heart sub-structures considered here is DeepLabv3+. It leverages the capabilities of atrous convolution that control the resolution of the feature map obtained from a deep convolutional neural network and multi-scale information adjusting filters' field of view.

The performance of DeepLabv3+ is evaluated using different loss functions during training and computing segmentation performance metrics on the test data. The results show that performance sensitive metrics based loss functions perform better as compared to pixel based loss functions.

Automated Cell Phenotyping in Spatial Proteomics via LLM-Guided Knowledge Distillation

Aroj Hada[1,2], Denis Schapiro[1,2,3], Jovan Tanevski[1,2]

[1] *Institute for Computational Biomedicine, Faculty of Medicine, Heidelberg University and Heidelberg University Hospital, Heidelberg, Germany*
[2] *Translational Spatial Profiling Center (TSPC), Heidelberg, Germany*
[3] *Institute of Pathology, Faculty of Medicine, Heidelberg University and Heidelberg University Hospital, Heidelberg, Germany*

Highly multiplexed imaging enables antibody-based spatial proteomics, revealing the spatial organization and localization of distinct cell populations at a single cell resolution. One of the first and central tasks in the analysis of spatial proteomics data involves inferring cell types from the data, i.e. cell phenotyping, a complex task requiring extensive manual interventions. Although many automated / semi-automated methods have been proposed, the challenges are rooted in i) the diversity of marker combinations and cell types encountered during inference, ii) the need for expert knowledge for accurate annotation, and iii) the lack of generalization to novel datasets as a result of inherent heterogeneity within the specific tissue and condition.

Here, we propose an agentic AI framework for automated phenotyping, leveraging the ability of Large Language Models (LLMs) as domain experts to address these challenges. To this extent, we provide domain-specific knowledge to the LLM through Retrieval Augmented Generation (RAG) from published literature on cell type and marker relationships. During inference, our approach involves two steps to map expert LLM knowledge to data comprising intensities of measured markers extracted for each cell from multiplexed images. First, given a list of markers, we distill a 'decision matrix' to map marker importances to cell-types and use it to make initial annotations. Second, the LLM revises annotations

where a cell has either a low probability of assignment to a cell type or equally high probability to be assigned to multiple cell types. Finally, a tree-based classifier is trained on the high confidence annotations from both steps to predict the cell type for the remaining cells. Experimental results show that our approach achieves a state-of-the-art classification performance comparable to related methods across multiple expert-annotated datasets.

Comparative Framework for Integrating Digital Pathology and Spatial Transcriptomics

Sebastián González Tirado[1,2], Jovan Tanevski[1,2]

[1] *Institute for Computational Biomedicine, Faculty of Medicine, Heidelberg University and Heidelberg University Hospital, Heidelberg, Germany*
[2] *Translational Spatial Profiling Center (TSPC), Heidelberg, Germany*

The convergence of digital pathology and spatial omics has generated significant interest due to its potential to improve predictive performance in clinically relevant tasks, guide cohort and panel design, and reduce costs. At this intersection, we present a systematic approach to explore and quantify the mapping performance between features extracted from H&E-stained tissue sections and high-dimensional molecular profiles obtained from spatial transcriptomics and the underlying biological processes they represent.

Evaluating the Clinical Utility of Pre-/Post-Operative Brain Tumor Volume Estimation using SwinUNETR

Simon Bele[1], Dejan Georgiev[1,2], Aleksander Sadikov[1]

[1] *University of Ljubljana, Faculty of Computer and Information Science*
[2] *University Medical Centre Ljubljana, Department of Neurology, Ljubljana*

Glioblastoma multiforme (GBM) is a highly aggressive brain malignancy where surgical resection remains a cornerstone of treatment. Accurate estimation of tumor volume before and after surgery is clinically critical, both for surgical planning and for assessing residual disease as a predictor of outcome. In this study, we evaluated the clinical utility of SwinUNETR, a transformer-based medical image segmentation model, for pre- and post-operative tumor volume estimation using the BraTS 2025 dataset. The model was trained and validated on four MRI modalities, with performance assessed using Dice score and voxel-level volume error. Results showed robust segmentation of whole tumor (WT) regions (Dice ≈ 0.89) and moderate performance for tumor core (TC) (Dice ≈ 0.75), with systematic underestimation of tumor volumes on average in both pre- and post-operative cases. These findings suggested that SwinUNETR could provide clinically relevant volumetric estimates when the size of the tumor was sufficiently large, though further refinement was required to address biases and improve reliability.

TT-XAI: Trustworthy Clinical Text Explanations via Keyword Distillation and LLM Reasoning

Kristian Miok[1], Blaž Škrlj[2], Daniela Zaharie[3], Marko Robnik Šikonja[1]

[1] *University of Ljubljana, Faculty of Computer and Information Science, Slovenia*
[2] *Jožef Stefan Institute, Slovenia*
[3] *West University of Timisoara, Romania*

Trustworthy AI in clinical text processing is critical for the safe adoption of decision support tools. Clinical language models often struggle with long, noisy electronic health records (EHRs), producing opaque predictions and low-quality explanations. We propose TT-XAI, a lightweight framework that improves both classification performance and interpretability through keyword-based distillation and large language model (LLM) reasoning. Discharge notes are distilled into domain-relevant keyword representations to boost BERT classifier performance and produce more faithful explanations via a focused LIME variant. Furthermore, keyword-guided prompts elicit concise clinical reasoning from LLMs, enhancing clarity and clinical relevance. Evaluation on kidney stone cases in MIMIC-IV shows consistent gains in predictive performance, deletion-based explanation fidelity, and both LLM and human expert ratings of reasoning quality. TT-XAI demonstrates a scalable pathway toward more trustworthy, auditable AI in clinical NLP.

Transparent and Reproducible Benchmarking of Machine Learning Methods in Stroke Data

Dimitar Trajkov[1], Ana Kostovska[1], Panče Panov[1], Dragi Kocev[1]

[1] *Jožef Stefan Institute*

Predicting brain stroke outcomes using machine learning is an important task with potentially high clinical impact. Yet, there is a lack of benchmarking studies that investigate which methods perform best on stroke-related data. Moreover, reproducibility remains a major challenge, as many studies lack transparent and standardized reporting practices. To address this, we conduct a comprehensive benchmark evaluating 12 classification and 27 regression methods across 8 publicly available datasets. Our analysis shows that dataset characteristics often influence performance more than the choice of the method. In addition, to improve trustworthiness, we introduce a semantic, ontology-based framework that formally describes the entire benchmarking pipeline, including data preprocessing, model configuration, and performance evaluation metrics. All results are published as a knowledge graph, accessible via a public querying endpoint and an interactive online catalogue for easy exploration and reuse.

Evaluation of Knowledge Graph Construction Methods for the Stroke Domain

Elena Atanasoska[1], Boshko Koloski[1], Dragi Kocev[1]

[1] *Jožef Stefan Institute*

Knowledge graphs (KGs) provide a powerful framework for representing complex, interrelated data—an essential capability for navigating intricate domains such as medicine. This is particularly true in the study of stroke, one of the leading causes of death and long-term disability worldwide. Currently, there is no comprehensive KG for the stroke domain, largely because the specialized knowledge embedded in the vast medical literature makes manual curation prohibitively time-consuming. In this work, we explore automated construction of a Stroke Knowledge Graph by comparing five relation-extraction methods and proposing novel *LLM-as-a-judge*-based evaluation. Specifically, we benchmark one unsupervised system (OpenIE), two supervised frameworks (REBEL and ReLiK), and two large-language-model approaches (Gemma 2 9B and Gemma 2 27B). Our contributions are twofold: first, we provide a systematic evaluation of multiple extraction strategies tailored to a domain-specific KG; second, we propose a hybrid evaluation protocol—combining traditional statistical metrics with an LLM-as-judge paradigm—to assess graph quality more holistically. Our results demonstrate that LLM-based methods hold particular promise for

generating a robust, high-coverage Stroke Knowledge Graph, a key resource for accelerating both research and clinical decision-making in stroke care.

Priming Deep Neural Networks with Knowledge Graphs Improves the Explainability of Condition-specific Gene Sets in Arabidopsis

Aleš Kert[1], Vid Modic[2], Tomaž Curk[1], Carissa Bleker[2], Jan Zrimec[2]

[1] *Bioinformatics laboratory, Faculty of Computer and Information Science*
[1] *Department of Biotechnology and Systems Biology, National Institute of Biology*

An important prerequisite to understanding how a plant functions and responds to the environment is to determine which gene expression patterns are associated with a specific tissue type and external perturbation response. Neural network-based methods can provide subsets of highly informative genes for such predictions using backpropagation-like methods. Here, we propose that integrating prior molecular knowledge related to gene expression within neural network architectures can lead to more reliable and insightful models, improving identification of tissue and perturbation-related gene sets. The Comprehensive Knowledge Network used to this end contains information on protein-DNA and protein-protein interactions. We first construct an Arabidopsis tissue- and perturbation-specific gene expression resource from published datasets and metadata. We then develop a pipeline comprising batch effect correction, prediction training, and model explanation. To address batch effects, we implement and evaluate several approaches, finding that Conditional Variational Autoencoders achieve the highest performance among tissue types, outperforming the other methods. We develop and train deep neural networks models to classify the underlying tissue types and perturbation groups using gene expression patterns as input. The knowledge graph-based models incorporating the prior molecular knowledge as additional network layers achieve similar classification performance as baseline models. However, the analysis of model explainability, by computing class specific relevance scores per gene, demonstrates that the knowledge graph-based models outperform baseline models by prioritising biologically relevant genes, known to be related to specific tissue types and molecular processes (e.g. phytohormone and stress-related responses). Our results thus demonstrate the applicability and reliability of knowledge graph-primed deep learning for identifying condition-specific genes and gene sets.

Context-Aware Predictor of T-cell Antigens (CAPTAn) Uncovers Novel Immunogenic Microbial and Viral Peptides

Martin Stražar[1], Jihye Park[2], Jennifer Abelin[2], Hannah Taylor[2], Thomas Pedersen[2], Damian Plichta[2], Yuan-Mao Hung[2], Eric Brown[2], Basak Eraslan[2], Kayla Ortiz[2], Karl Clauser[2], Steven Carr[2], Ramnik J. Xavier[3], Daniel Graham[2]

[1] *National Institute of Biology*
[2] *Broad Institute of MIT and Harvard*
[3] *Massachusetts General Hospital*

Microbes in the gut promote education of the adaptive immune system through recognition of peptide antigens recognized by antigen-presenting cells (APCs) and T cells. Here, we retrieve hundreds of thousands of unique peptides presented by engineered antigen-presenting cells. This expansive dataset revealed novel binding rules and properties of naturally presented peptides. Using neural networks and ensemble learning, we develop Con-text-Aware Predictor of T cell Antigens (CAPTAn). Improving on existing models, we predict peptide ligands from their whole protein sequence, increasing the accuracy of retrieval from large proteomes. Our method leverages both amino acid binding preferences

and contextual features related to antigen structure, trafficking, and localization, which are learned in annotation-agnostic manner. Integration of machine learning, microbial genomics and metatranscriptomics predicts prevalent antigens from the human micro-biome and SARS-CoV-2 which are experimentally validated through cytokine assays and clonally expanded TCRs from CD4+ T cells. Exposing features of antigenicity through CAPTAn enabled discovery of bacterial and viral antigens that drive functionally heterogeneous T cell responses in humans. CAPTAn is freely available at https://broad.io/captan.

Multi-Objective Active Learning for Nanobody Development

Katharina Dost[1], Klara Kropivšek[2], Christian L. Camacho Villalón[1], Sašo Džeroski[1], Ario de Marco[2]

[1] *Jožef Stefan Institute*
[2] *University of Nova Gorica*

Designing nanobodies with desirable biological and biophysical properties is experimentally costly and time-consuming. We present a multi-objective active learning framework that guides nanobody selection by jointly optimizing for model informativeness, expression yield, and target binding affinity. By combining predictive modeling with evolutionary multi-objective optimization, our approach proposes candidates that are both experimentally viable and strategically informative for model improvement. This framework reduces costs to obtain property information, enhances model generalization, and supports efficient nanobody discovery and ease of development. While currently tailored to nanobody engineering, the method generalizes to other domains requiring data-efficient multi-objective optimization.

Assessing the Potential of Large Language Models for In-Context Few-Shot Molecular Learning in Ligand-Based Drug Design

Marko Jukič[1], Anja Kolarič[1], Urban Bren[1]

[1] *Faculty of Chemistry and Chemical Technology, University of Maribor*

Large language models trained on chemical and natural language corpora are emerging as generative engines for molecular discovery. Unlike reinforcement learning or enumerative fragment-based approaches, LLMs can perform zero- or few-shot in-context learning, enabling property-guided molecule design without model retraining. In this study, we evaluate the capacity of general-purpose models to generate novel small molecules relevant to SARS-CoV-2 antibacterial design on NSP14, 3CLpro, PLpro, and neuropilin targets. We benchmark LLM-driven generation against established frameworks (like REINVENT), and fragment-based enumerations such as FASMIFRA and SAVI SPACE. We tested both zero-shot prompting and few-shot experiments incorporating example SMILES-target pairs. Outputs were analyzed for validity and diversity, and evaluated by ligand-based similarity scoring and docking against target structures. Results indicate that few-shot prompting improves SMILES validity and scaffold diversity and that this general approach successfully parallels within uniquely generated structures to full enumerated libraries. Our findings highlight the practical potential of LLMs as complementary generative tools in ligand-based drug design.

Deep learning-Driven Synthetic Data Generation for Secure Ear Biometrics

Iyyakutti Ganapathi[1], Darian Tomašević[1], Peter Peer[1]

[1] *University of Lubljana*

Generative Adversarial Networks (GANs) have emerged as powerful tools for synthesizing high-fidelity biometric imagery, yet their tendency to memorize training samples raises serious privacy concerns. In this work, we revisit StyleGAN3 for ear biometrics and introduce a frequency-gated variant designed to enhance spectral diversity during generation. While the model successfully produces visually convincing ear images, we empirically demonstrate that it exhibits strong identity memorization: a standard membership inference attack achieves an AUC of 0.87, revealing substantial leakage of training identities. Our analysis uncovers that StyleGAN3's alias-free architecture, even when augmented with frequency regularization, continues to overfit under limited-domain biometric data. We further benchmark privacy leakage across synthesis conditions and show that generated samples often remain within the distributional neighborhoods of training images, limiting their utility for privacy-preserving augmentation. These findings highlight a critical gap between visual quality and privacy robustness in current GAN-based biometric synthesis pipelines. We conclude by motivating the transition toward diffusion-based and hybrid identity-conditioned frameworks, where frequency gating and identity-mixing strategies may serve as effective safeguards against memorization. Our study provides the first systematic privacy evaluation of StyleGAN3 in a biometric setting and establishes baselines for future research in secure synthetic data generation.

AI & Environmental Science

Track Chairs

Matjaž Ličer, Slovenian environment Agency, National Institute of Biology - SMASH supervisor
David Kocman, Jožef Štefan Institute
Nataša Atanasova, University of Ljubljana

Monday - AI for Climate, Weather, and Environmental Systems

14:30 - 15:15	**Alexander Barth** Generative Deep Learning for Satellite Data Reconstruction
15:15 - 15:35	Discovery of Entity Types for Climate Change
15:35 - 16:00	An AI-Enhanced GIS Platform for Natural Language-Based Environment Analysis by Non-GIS-Specialists
16:00 - 16:30	**Coffee Break**
16:30 - 16:50	Forecasting high-frequency sea-level oscillations in the Adriatic Sea using deep neural networks
16:50 - 17:10	A neural network-based observation operator for weather radar data assimilation
17:10 - 17:30	AI-Based Correction of Ocean Surface Wind Biases in Global Numerical Weather Prediction
17:30 - 17:50	Escaping the Plausiblity Trap: Verified Agentic Pipelines for Climate-Scale Nonlinearity

Tuesday – AI for Ecology, Biodiversity, and Agriculture

09:00 - 09:45	**Joaquim Comas Matas** Integrating AI in Urban Water Management: Intelligent Decision Support Systems for the Twin Transition
09:45 - 10:10	Multi-objective Optimization of Nature-based Solutions
10:10 - 10:30	ML-based Turbidity forecasting in Drinking Water Treatment Plants for improved treatment
10:30 - 11:00	**Coffee Break**
11:00 - 11:20	Modelling Yield of Diverse Potato Varieties from Multi-Season Field Trial Data
11:20 - 11:40	Classification of Woody Vegetation Landscape Features from National Orthophoto
11:40 - 12:00	Predator detection in low-resolution thermal videos: Insights from Analyzing the New Zealand Wildlife Thermal Imaging dataset
12:00 - 12:20	Spatial approach for assessing vulnerability to urban flooding: a proposal for a multidimensional index
13:00 - 14:30	**Lunch Break**

14:30 - 14:50	Forecasting Chlorophyll-a Concentrations in the Gulf of Trieste with Machine Learning Techniques
14:50 - 15:10	Deep Learning Approach to Landmarking and Origin Classification for Gilthead Seabream (Sparus aurata)
15:10 - 15:30	Automated modelling of Lake Bled, Slovenia
15:30 - 15:50	Modeling Fungal Diversity in Coastal Sands with Interpretable Machine Learning
16:00 - 16:30	**Coffee Break**
16:30 - 16:50	Sparsity introduction in Bayesian Autocorrelation Matrix factorization for organic aerosol source apportionment
16:50 - 17:10	ML analysis for absorption measurements correction schemes – A test study
17:10 - 17:30	Assisting organic aerosol source apportionment with a multinomial logistic regression classifier
17:30 - 17:50	Graph Neural Network (GNN) powered electronic structure analysis of CuO–SnO_2 heterostructure: towards AI-driven H_2S gas sensor optimization

Generative Deep Learning for Satellite Data Reconstruction

Alexander Barth[1]

[1] *University of Liège*

One of the major strengths of satellite data is its ability to provide near-global coverage. However, satellites operating in the visible or infrared range are affected by clouds (among other factors), which can significantly reduce the spatial and temporal coverage.

For many applications, complete coverage is either desired or even necessary. In addition, it is increasingly important to accurately reflect the underlying uncertainty of the reconstructed observations. In this study, we show the use of neural networks to fulfil these goals. We present DINCAE (Data-INterpolating Convolutional Auto-Encoder), a neural network that can be trained on incomplete images and is able to provide an error estimate of the reconstructed field. The reconstructed data and their associated error estimates are validated with in situ observations and withheld satellite data. DINCAE has been applied to various parameters, including sea surface temperature, chlorophyll-a concentration, total suspended matter and along-track sea-level altimetry.

Generative deep learning techniques (denoising diffusion models and conditional flow matching) can naturally provide an ensemble of reconstructions, where each member is spatially coherent with the scales of variability and with the available data. Rather than providing a single reconstruction, an ensemble of possible reconstructions can be generated, and the ensemble spread reflects the underlying uncertainty. We show how this method can be trained from a collection of satellite data without requiring prior interpolation of missing data. The reconstruction method is tested with chlorophyll-a concentration from the Ocean and Land Colour Instrument (OLCI) sensor aboard the Sentinel-3A and Sentinel-3B satellites.

The spatial scales of the reconstructed data are assessed via a variogram, and the accuracy and statistical validity of the reconstructed ensemble are quantified using the continuous ranked probability score and its decomposition into reliability, resolution, and uncertainty.

Discovery of Entity Types for Climate Change

Andrija Poleksić[1]

[1] *Faculty of Informatics and Digital Technologies*

Named Entity Recognition (NER) is a fundamental task in information extraction, yet general-purpose NER categories often fail to capture the specificity required for specialized domains such as climate change research. We propose automatic construction of a domain-specific NER type set with minimal supervision, leveraging a schema-based bottom-up approach to knowledge graph construction. The process begins with the identification of 655 core climate change-related terms, sourced from authoritative domain-specific resources. These terms are then semi-automatically aligned with Wikidata using SPARQL queries to take advantage of its hierarchical structure. A neighbourhood graph is constructed based on instance of (P31) and subclass of (P279) properties, forming the basis for community detection via the weighted Louvain algorithm. The resulting 59 communities are manually analyzed to derive a final set of 21 NER types specific to Climate Change domain. Validation against existing ontologies and terminological knowledge base (SWEET, ENVO, and EcoLexicon) reveals that the SWEET ontology provides the highest coverage, containing 57.25% of core terms and 65.38% of the proposed NER types.

An AI-Enhanced GIS Platform for Natural Language-Based Environment Analysis by Non-GIS-Specialists

Vid Primožič[1], Vid Jakopin[1], Jaka Kravanja[1]

[1] *Flycom Technologies d.o.o*

Geospatial data plays an essential role in environment sciences, land administration, planning, and Earth observation. However, its complex structure makes it difficult for non-experts to access or analyze information from various sources. This complexity often creates a barrier to wider use, limiting the potential of geospatial data. To overcome this barrier, we developed a prototype app that uses a Large Language Model (LLM) to turn natural language questions into geospatial operations. It connects to a PostGIS database and follows a step-by-step process to understand the request, find the right data, and run the needed SQL queries automatically. Our system follows a structured and modular design. The AI agent creates a plan, finds the right spatial data, and step-by-step generates and runs queries in a controlled loop. We tested it on real vector and raster data from Slovenia, using both open and official geodetic sources. The approach worked well across many types of spatial tasks and can be used with different language models, making it a promising tool for easier geospatial analysis. By lowering barriers to geospatial analysis, the system enables broader use of AI for environmental data and more timely, informed decisions in land and resource management.

Forecasting high-frequency sea-level oscillations in the Adriatic Sea using deep neural networks

Iva Međugorac[1], Nikola Metličić[2], Marko Rus[3], Jadranka Šepić[2], Matej Kristan[4], Matjaž Ličer[5]

[1] *University of Nova Gorica*
[2] *Faculty of Science, University of Split, Croatia*
[3] *Slovenian Environment Agency, Ljubljana, Slovenia*
[4] *Faculty of Computer and Information Science, University of Ljubljana, Slovenia*
[5] *National Institute of Biology, Ljubljana, Slovenia*

Meteotsunamis represent rare occurrences of extreme high-frequency sea level oscillations (HFOs) that occasionally affect open coasts, ports and harbors. These phenomena are triggered by spatially confined atmospheric disturbances and form only when certain characteristics of these disturbances enable resonant energy transfer between the atmospheric forcing and long ocean waves, whose properties depend on the bathymetric features of the coastal shelf in front of open coast, bays and harbors. Several locations in the Mediterranean are particularly vulnerable to meteotsunamis.

In this study, we introduce the first application of deep learning methods for predicting the height of HFOs at the Adriatic tide-gauge station Bakar. Although Bakar is not particularly exposed to meteotsunamis, its long-term high-frequency sea level measurements (from 2003) make it suitable for model development. The pretrained models developed for this location can be transferred and further fine-tuned for application at other Adriatic sites, including those with shorter data records and greater meteotsunami susceptibility. We trained deep convolutional neural networks on observed sea levels combined with atmospheric reanalysis fields (ERA5). A large number of experiments was carried out to evaluate various network designs, input combinations, and prediction targets to identify the most effective configuration.

The main conclusions are: (i) the model is capable of reasonably forecasting daily maximum HFO amplitudes up to three days in advance; (ii) forecasts are more accurate for low-amplitude HFOs; (iii) higher-amplitude events are typically underestimated; (iv) expanding the input data (for example, by increasing the time window or adding extra sea-level components) does not improve the forecast quality; (v) for 1-minute sea-level predictions,

while the model notably underestimates amplitudes at daily forecast scales, it performs much better over shorter horizons (e.g., next six hours).

A neural network-based observation operator for weather radar data assimilation

Marco Stefanelli[1], Ziga Zaplotnik[2], Gregor Skok[1]

[1] *UNI-LJ*
[2] *ECMWF*

Forecasting convective storms is a major challenge in Numerical Weather Prediction (NWP). Data Assimilation (DA) improves the initial condition and subsequent forecasts by combining observations and previous model forecasts (background). Weather radar provides a dense source of observations in storm monitoring. Therefore, assimilating radar data should significantly improve storm forecasting skills. However, the short-range precipitation forecast performed by extrapolating rainfall patterns (nowcasting) from radar data is often better than numerical model-based forecasting with DA (Fabry and Meunier, 2020). This is related to the fact that the radar data only provides information on the precipitation pattern and intensity in the area affected by the storm. Furthermore, it does not directly provide information on the environmental conditions of the storm, such as temperature, wind, and humidity, neither within the precipitation region nor in the areas far from the storm. A potential solution involves using machine learning (ML) to construct the DA observation operator to generate a model-equivalent of the radar data. In this approach, NWP model fields (temperature, wind components, relative humidity) would serve as input, and radar observations would be the output of an encoder-decoder neural network. The constructed observation operator describes a non-linear relationship between the NWP model storm-related variables and radar observations, while its Jacobian allows radar information to infer other variables in the data assimilation, potentially enhancing storm forecasting skills.

AI-Based Correction of Ocean Surface Wind Biases in Global Numerical Weather Prediction

Evgeniia Makarova[1], Marcos Portabella[1], Ad Stoffelen[2], Veronica Vilaplana[3], Ad StoffelenManuel García-León[4], Marcos Sotillo[4]

[1] *Institute of Marine Sciences*
[2] *KNMI*
[3] *UPC*
[4] *NOW*

Global Numerical Weather Prediction (NWP) model sea-surface winds are commonly used to force ocean models due to their continuity in time and space. However, these outputs exhibit local biases, with one of the most persistent and systematic biases occurring in sea surface wind direction. One key source of observational wind data is the Advanced Scatterometer (ASCAT), a C-band radar instrument aboard the MetOp satellite series, which provides high-quality wind measurements near the sea surface. These observations are assimilated into both the ECMWF operational forecasts and the ERA5 reanalysis; however, systematic biases in the wind fields remain.

A previous approach by Trindade corrects these biases using time-averaged differences between scatterometer and model winds, a method operationally implemented in Copernicus Marine Data Store (MDS) products for both ECMWF operational forecasts and

reanalysis. Nevertheless, its performance degrades during periods of sparse scatterometer coverage.

To overcome these limitations, we propose a machine learning (ML) approach that predicts and corrects the wind stress biases using NWP-based variables as inputs, including 10-m stress-equivalent winds, sea level pressure, surface temperature, humidity, sea surface temperature, and ocean surface currents.

Several architectures were evaluated, including XGBoost, simple feed-forward neural networks, and U-Net. The feed-forward networks demonstrated the most robust performance, reducing global error variance by 14% when trained on five years of data. In contrast, the U-Net architecture achieved lower RMSE on a smaller training subset but led to a noticeable drop in spatial variance.

Future work will focus on implementing and evaluating architectures, potentially generative neural networks, that balance spatial variance and classical metrics like RMSE, while also incorporating physical constraints.

Escaping the Plausiblity Trap: Verified Agentic Pipelines for Climate-Scale Nonlinearity

David Lewis[1], Enrique Zueco[1]

[1] *AIXC*

Autonomous "AI Scientists" can produce results that read as rigorous yet quietly violate physical laws, a failure mode we call the plausibility trap. In climate science—marked by nonlinear dynamics and multi-scale couplings—small violations can cascade into high-stakes policy errors. We present a conservative, multi-stage Verified Agentic Pipeline (VAP) that couples physics-informed learning, agentic retrieval and planning, formal specification checking, uncertainty quantification, and risk-informed governance. We enforce physical consistency via PINNs and neural operators hybridized with data assimilation (4D-Var/EnKF); structure reasoning with agentic literature tools; gate claims through specification verification and coverage-guaranteed uncertainty (e.g., conformal prediction); and align evidence with NIST AI RMF, ASME V&V 40, and NASA-STD-7009B while mapping geoengineering to the Oxford Principles, LC/LP, CBD, and National Academies guidance. The result preserves creative search while preventing propagation of physically invalid findings in climate applications.

Integrating AI in Urban Water Management: Intelligent Decision Support Systems for the Twin Transition

Joaquim Comas Matas[1]

[1] *University of Girona*

The development and application of intelligent multicriteria decision support systems (iDSS) for the sustainable management of the urban water cycle has been one of Dr. Joaquim Comas's core research lines since the beginning of his PhD in 1995. With nearly three decades of experience, he has become a recognized expert in the integration of mathematical modelling, statistical analysis, and artificial intelligence techniques to support planning, design, simulation, and advanced control of urban water infrastructures.

His early research focused on wastewater treatment systems, mainly activated sludge, progressively expanding to address the integrated management of the entire urban water cycle, including supply, collection, treatment, and the interaction with receiving water bodies. He has in-depth expertise in knowledge-based systems, such as rule-based reasoning and case-based reasoning, as well as machine learning techniques. Approximately one third

of his journal publications fall within the field of computer science, with a strong emphasis on artificial intelligence applied to environmental systems.

In the mid-2000s, Dr. Comas initiated a pioneering research line on the intelligent control of membrane bioreactors (MBR). This work has since evolved to include the optimisation of cutting-edge membrane technologies, such as integrated MBRs with reverse osmosis or nanofiltration units, forward osmosis systems, and membrane recycling, all aligned with the principles of the circular economy.

Today, his main research activity continues to revolve around the development of intelligent iDSS tools to promote circularity in the urban water cycle, particularly through water reuse and nature-based systems. His work has not only advanced the scientific field but also had practical impact: in 2003, he founded one of the first spin-off companies of the University of Girona, focused on developing intelligent control systems for wastewater treatment plants (nowadays https://www.createch360.com/).

Multi-objective Optimization of Nature-based Solutions

Shengnan Yang[1], Matej Radinja[1], Nataša Atanasova[1]

[1] *University of Ljubljana*

Urbanization and climate change have intensified urban challenges, including flooding, water pollution, and urban heat islands. Nature-based solutions (NBS) offer a sustainable, cost-effective approach to address these issues while delivering environmental, social, and economic co-benefits. This study aims to build a multi-objective optimization framework and derive Pareto-optimal NBS scenarios via the multi-objective optimization and assessment model. The methodology was applied to Tivoli Park in Ljubljana, Slovenia, in which is a flood-prone area with mixed land uses. The hydrology-hydraulic model was built in Stormwater Management Model (SWMM) and calibrated using the monitored flow. The Non-Dominated Sorting Genetic Algorithm II (NSGA-II) takes new values for the areas of the NBS units and compares the hydrological response of the catchment with the baseline model simulation. The optimized results demonstrate that NBS scenarios are effective for flood reduction, peak flow control, pollutant reduction, water reuse, infiltration increase, evaporation increase, and green space increase. This research provides a deeper understanding and explanation of the trade-offs among NBS scenarios.

ML-based Turbidity forecasting in Drinking Water Treatment Plants for improved treatment

David Abert-Fernández[1], Nataša Atanasova[2], Hèctor Monclús[1], Daniel Kozelj[2]

[1] *University of Girona*
[2] *University of Ljubljana*

Rainfall induces peak events at the catchment of drinking water treatment plants. These peaks pose significant operational challenges, interfering with the treatment processes and affecting its performance. Although there are methods to smooth abrupt fluctuations in water quality, they often require the construction of infrastructure. However, these solutions may not be feasible in all contexts due to spatial constraints or cost. To address this gap, a turbidity forecasting algorithm was developed in this study. By forecasting the onset, intensity and duration of these events, treatment plants can implement preventive measures to maintain water quality and ensure operational efficiency. Strategies such as adjusting treatment parameters or pre-filling storage reservoirs prior to these events can help mitigate the impacts of peak turbidity, reducing stress on the system and optimize overall treatment performance. To address this task, this study used historical turbidity

and meteorological data in conjunction to machine learning models. Additionally, a conformal prediction approach was implemented to quantify the uncertainty associated with the developed models.

Spatial approach for assessing vulnerability to urban flooding: a proposal for a multidimensional index

Ana Noemí Gomez Vaca[1], Ignasi Rodríguez-Roda[1], Alexandra Popartan[1]

[1] *University of Girona*

This study introduces a methodology for evaluating vulnerability to urban flooding across different dimensions, by employing spatial data analysis. The methodology consists of four steps: (1) selection of indicators that reflect the vulnerability of an urban area, (2) normalization of the data for each selected indicator across all dimensions, (3) assignment of weights for each indicator and dimension, and (4) mapping and classification using spatial analysis, resulting in a grid. This study proposes a comprehensive list of 36 potential indicators for quantifying vulnerability, with each indicator falling under one of the four dimensions (social, economic, environmental, and physical) and three components of vulnerability (exposure, susceptibility, and resilience), which are part of Step 1. Additionally, the methodology is complemented by a data generation and spatial analysis technique for Step 4. The proposed methodology can serve planners and policymakers to make objective decisions, based on vulnerability quantification through index for long-term vulnerability to flooding, considering each dimension separately, as well as integrating with each other, using a multidimensional and spatial representation of flood risk vulnerability.

Modelling Yield of Diverse Potato Varieties from Multi-Season Field Trial Data

Angelika Vižintin[1], Eva Turk[1], Maja Zagorščak[1], Alexandra Ribarits[2], Bianca Doevendans[3], Guus Heselmans[3], Ivana Imerovski[4], Jeroen Bakker[4], Pieter-Jelte Lindenbergh[4], Jan Zrimec[1,5], Kristina Gruden[1],

[1] *National Institute of Biology*
[2] *Austrian Agency for Health and Food Safety*
[3] *Meijer Potato*
[4] *HZPC*
[5] *University of Ljubljana, Biotechnical Faculty*

Potato (Solanum tuberosum L.) is the world's fourth most important crop, and accurate yield prediction is crucial for optimizing management, resource use, and food security. For this study, potato field trials were conducted in three European countries between spring 2021 and autumn 2023, cultivating 44 diverse varieties, under irrigated and non-irrigated conditions. To monitor plant growth and development in relation to environmental conditions, extensive measurements were collected throughout the growing seasons. These included sensor-based monitoring of soil and air temperature and moisture, as well as drone-based observations to assess vegetation cover and key vegetation indices, such as the normalized difference vegetation index (NDVI). After harvest, yield per hectare was recorded. The field trial data were further supplemented with data from nearby weather stations. Kernel ridge regression models were developed to predict yield per hectare. By training models on data collected up to different days post planting, we found that reliable yield predictions are possible as early as the first two months of the growing season. To identify the most informative features, we applied recursive feature elimination followed by testing combinations of the top-ranked features. The most important predictors included

environmental factors (soil moisture, humidity, air temperature), consistent with agricultural knowledge about heat stress effects, as well as drone-derived features, which enable plot-level predictions. Our results demonstrate that the developed models can effectively predict the yield of new potato varieties—varieties that were not included in model tuning, training, or feature selection—and highlight key environmental variables to monitor during the growing season that have the greatest impact on yield. These in-sights can guide both data collection strategies and agronomic practices aimed at optimizing potato production.

Classification of Woody Vegetation Landscape Features from National Orthophoto

Adam Gabrič[1], Dejan Grigillo[2], Žiga Kokalj[1]

[1] *Research Centre of the Slovenian Academy of Sciences and Arts, Institute of Anthropological and Spatial Studies*
[2] *University of Ljubljana, Faculty of Civil and Geodetic Engineering*

Woody vegetation landscape features (single trees, trees in rows, groups of trees and bushes, hedgerows, and riparian vegetation) provide valuable ecosystem services. However, data on their locations and extent remains lacking. We tested their classification from national orthophoto using HRNet convolutional neural network. The classifications reach overall accuracies over 90% showing that the model accurately predicts the background pixels. Other calculated accuracies (Jaccard index, user's and producer's accuracy) are lower proving that the high overall accuracies are consequences of class imbalance in the reference data. For most of the classifications the false positives seem to be more prominent than false negatives. Most of the classes reach JI 14—19% with the exceptions being trees in rows and riparian vegetation. Trees in rows have lower JI which is most likely caused by smaller number of patches included in the training of the model. On the other hand, the classification of riparian vegetation reaches highest overall JI (34.17%).

Predator detection in low-resolution thermal videos: Insights from Analyzing the New Zealand Wildlife Thermal Imaging dataset

Katerina Taskova[1]

[1] *The University of Auckland*

Despite significant progress in automated wildlife monitoring using thermal imaging in various ecological and conservation contexts, several gaps remain unaddressed in the literature. Existing methods predominantly rely on single-image detection and classification, which often fail to capture the temporal dynamics of animal movements, particularly relevant for scenarios involving overlapping or fast-moving animals. The lack of temporal information limits the ability to track individuals across video frames, an essential feature for robust monitoring systems. Including temporal information could enhance detection accuracy while providing a framework for effectively leveraging sequential data. Furthermore, achieving high accuracy with lightweight models optimized for low-resolution thermal imagery remains a key challenge. This is particularly important for processing low-resolution imagery, where visual information is already limited in spatial resolution and texture. Moreover, existing approaches often overlook the importance of frame selection strategies, which are critical in real-world applications where not all frames contain equally informative or detectable animal instances. Finally, analyzed datasets are relatively small in size, context-specific, predominately airborne (collected with drones) that are not useful for small-sized predators like rats and mice, and typically not open-source. By bridging these gaps, this study aims to offer practical and scalable solutions for real-time wildlife

monitoring in diverse ecological settings. We developed and analyzed a light-weight novel method for species classification in low-resolution thermal videos, based on the open-source New Zealand Wildlife Thermal Imaging dataset (collected by the Cacophony project).

Forecasting Chlorophyll-a Concentrations in the Gulf of Trieste with Machine Learning Techniques

Katja Bengeri[1], Jana Faganeli Pucer[1], Martin Vodopivec[2]

[1] *University of Ljubljana*
[2] *National Institute of Biology*

Phytoplankton represents an important indicator of the ecological state of the marine environment. As in-situ measurements are difficult to obtain and time-consuming to analyze, researchers are exploring alternative methods to estimate chlorophyll-a concentrations from other environmental parameters. In this study, we investigated the feasibility of estimating chlorophyll-a concentrations in the Gulf of Trieste using machine learning models. A set of ecologically relevant features was selected in collaboration with domain experts to construct the input dataset. Since real chlorophyll-a measurements are scarce, we employed a transfer learning strategy: we pre-trained a model on satellite-derived chlorophyll-a data and fine-tuned it with ground measurements. For this purpose, multilayer perceptrons (MLPs) and convolutional neural networks (CNNs) were applied. The application of transfer learning led to an evident improvement in predictive performance, demonstrating its potential to enhance chlorophyll-a estimation in data-limited marine environments.

Deep Learning Approach to Landmarking and Origin Classification for Gilthead Seabream (Sparus aurata)

Josip Šarić[1], Igor Talijančić[2], Josip Zavada[3], Luka Žuvić[2], Siniša Šegvić[3], Tanja Šegvić Bubić[3]

[1] *University of Ljubljana Faculty of Computer and Information Science*
[2] *Institute of Oceanography and Fisheries*
[3] *University of Zagreb, Faculty of Electrical Engineering and Computing*

Geometric morphometrics has revolutionised the study of morphological structures in biological research by providing precise tools for shape analysis. However, manual digitisation of landmarks remains prone to variation and introduces measurement error that can obscure or distort the biological signal. This study addresses these challenges by proposing a novel deep learning method for automatic landmark placement in gilthead seabream, a key species in Mediterranean aquaculture. Using the collected dataset of 2052 specimens of wild, farm-associated and farmed origin, the model achieved sub-millimetre test accuracy with an average Euclidean distance of 0.94 mm. The model surpassed the novice human annotator and achieved performance comparable to an expert. Systematic and random measurement errors were quantified using advanced statistical techniques, demonstrating that the method significantly reduces interoperator variability. Semilandmarks improved the measurement reliability by reducing systematic bias and helped to stabilise the placement of landmarks in regions of high morphological complexity. Analysis of landmark configurations revealed that configurations with included semilandmarks improved the detection of the biological signal while reducing measurement error. The results underline the potential of integrating deep learning into geometric morphometrics workflows to improve the reproducibility and scalability of shape analyses. The proposed integrated approach is a step towards a cost-effective solution for important challenges in aquaculture management, such as escapees detection in wild populations and preventing fraud in fish origin labelling.

Automated modelling of Lake Bled, Slovenia)

Mateja Skerjanec[1], Laura Rant[1], Matej Radinja[1], Nataša Atanasova[1]

[1] *University of Ljubljana, Faculty of Civil and Geodetic Engineering*

Lake Bled, one of Slovenia's most visited natural landmarks, has experienced persistent algal blooms over the past century, despite multiple rehabilitation efforts. To support improved lake management, we conducted a national research project (V2-2370) focused on assessing the impact of pre-baiting fishing practices and other activities on lake water quality. An extensive database was compiled, combining historical data and new field measurements on nutrient loadings and the ecological and chemical status of the lake and its inflows.

Within the project, we developed a catchment-scale GWLF model that simulates hydrological and nutrient-leaching processes to estimate long-term nutrient loadings. These outputs were used as inputs for a lake nutrient-cycling model implemented in AQUASIM, assuming phosphorus as the primary limiting factor. Additionally, an automated modeling approach using the ProBMoT tool was tested, which applies a declarative formalism for describing the system. It utilizes a domain-specific modeling library to induce models from data. At first, this approach was only applied to the lake modelling part, using the already established aquatic ecosystem modeling library, compatible with ProBMoT. In the second stage, the proposed approach will also be applied to the catchment modelling part using the pre-established watershed modeling library.

Preliminary results indicate that the Mišca tributary is the dominant source of phosphorus and that internal lake loading significantly contributes to algal blooms. Future steps include exploring the advantages of automated modeling of Lake Bled within the AI4Sci Gravity project, in which we would like to connect both above-mentioned knowledge libraries. This integration will enable simultaneous automated modeling of nutrient loadings and lake water quality. The proposed AI-driven approach will support scenario testing and informed, model-based lake management strategies.

Modeling Fungal Diversity in Coastal Sands with Interpretable Machine Learning)

Monika Novak Babič[1], Katharina Dost[2], Ana Margarida Silva[2], Sašo Džeroski[1], João Brandão[2]

[1] *University of Ljubljana*
[2] *Jožef Stefan Institute*
[3] *National Institute of Health Dr. Ricardo Jorge*

Sandy beaches host diverse microbial communities, yet fungi in these environments remain underexplored despite their ecological roles and potential health implications. Some species tolerate extreme abiotic stressors such as intense solar radiation and fluctuating moisture, and previous work has documented antifungal resistance among isolates from recreational sands. These findings highlight the importance of understanding how climate and geography shape fungal distribution and diversity.

We present a large comparative dataset of fungal communities from 94 beaches across 17 countries, accompanied by detailed metadata on climate, geography, and site characteristics. Each sample includes descriptors such as sand type, urbanization level, season, and spatiotemporal context, along with meteorological variables including temperature, humidity, and sunshine hours.

To analyze these data, we employ tree-based machine learning methods that combine predictive performance with interpretability. Predictive clustering trees and ensemble approaches are applied to model species richness and the abundance of individual taxa as a

function of climatic and geographic factors. These models provide explicit, human-readable rules that clarify how environmental gradients and location influence fungal community structure.

This study builds directly on our earlier analysis of a Slovenian urban beach and extends the scope to a global scale. It demonstrates how interpretable AI methods can reveal large-scale ecological patterns in beach mycobiota and provide a framework for studying microbial diversity in relation to climate and geography.

Sparsity introduction in Bayesian Autocorrelation Matrix factorization for organic aerosol source apportionment)

Marta Via Gonzalez[1], Anton Rusanen[2], Jure Demšar[3], Yufang Hao[4], Manos Manousakas[5], Jianhui Jiang[6], Andrés Alastuey[7], Stuart K. Grange [8], Jean-Luc Jaffrezo[9], Vy Nguyen Thuy[9], Gaëlle Uzu[9], Griša Močnik[10], Kaspar R. Dällenbach[4]

[1] *Centre for Atmospheric Research - University of Nova Gorica*
[2] *Atmospheric Composition Research, Finnish Meteorological Institute*
[3] *Faculty of Computer and Information Science, University of Ljubljana*
[4] *Laboratory of Atmospheric Chemistry, Paul Scherrer Institute*
[5] *Environmental Radioactivity Aerosol Tech. for Atmospheric Climate Impacts, INRaSTES, National Centre of Scientific Research "Demokritos"*
[6] *School of Ecological and Environmental Sciences, East China Normal University*
[7] *Institute of Environmental Assessment and Water Research (IDAEA-CSIC)*
[8] *Climate and Environmental Physics, Physics Institute, University of Bern*
[9] *University of Grenoble Alpes, CNRS, INRAE, IRD, Grenoble INP, IGE*
[10] *Centre for Atmospheric Research - University of Nova Gorica*

The Positive Matrix Factorisation (PMF) algorithm (Paatero and Tapper, 1994) has been the most widely used receptor model for a long time and has only recently been challenged with new methodologies. The novel Bayesian auto-correlated matrix factorisation method (BAMF, Rusanen et al. 2024) integrates an auto-correlation term emulating real-world pollutant sources time evolution has produced higher accuracy compared to PMF. However, both PMF and BAMF struggle to provide well-separated profiles manifested as mixed time series contributions.

A sparsity-handling algorithm named horseshoe (HS) regularisation has been applied to BAMF in order to improve profile determination. The horseshoe application pushes some parameters to be close to zero and others to have large values (Piironen and Vehtari, 2017). The resulting profiles are expected to be less noisy and better representing the nature of the atmospheric pollution sources. Hence, the outcoming time series, which have been reported to be very sensitive to profiles adjustment, are expected to improve substantially.

Figure 1 shows the effect of BAMF+HS (in orange) compared to the regular BAMF (in blue) and the PMF (in green) on a toy dataset, consisting on an oversimplified dataset with very sparse profiles. The BAMF+HS results show contributions pushed to zero, making the profiles closer to the truth (in black) with respect to the less sparse results of BAMF and PMF.

This same comparison has been carried out on realistic synthetic datasets for 5 different sites to show the effectiveness of sparsity introduction into source apportionment. Four of these synthetic datasets were created from CAMx model outputs and reference profiles to make them challenging for the models, the other one was created from a more simplistic random walks approach. Preliminary results show better profile accuracy and a more balanced overall solution when the horseshoe prior is implemented in the model.

ML analysis for absorption measurements correction schemes – A test study)

Jesús Yus Díez[1], Jorge Perez[2], Lucas Alados-Arboleda[3], Luka Drinovec[1], Gloria Titos[3], Tuukka Petäjä[4], Andrés Alastuey[5], Xavier Querol[5], Griša Močnik[1]

[1] *University of Nova Gorica*
[2] *Nextail labs*
[3] *University of Granada*
[4] *University of Helsinki*
[5] *IDAEA-CSIC*

Light absorbing carbonaceous aerosols contribute to the Earth's atmospheric warming. To determine their contribution, measurements from across the globe are incorporated into climate models. The most used instruments for this measurement are filter-photometers (FP), which measure the attenuation of light through a filter where aerosols are deposited. Of all, the most deployed instrument across networks is the aethalometer AE33. More sophisticated FP, such as the Multi-Angle Absorption Photometer (MAAP) are also deployed in some stations.

FPs feature several artifacts, with cross-sensitivity to scattering being most important at high single scattering albedo, where error exceeds 100%. This cross-sensitivity to scattering has been found to have site-to-site variability and a correction scheme has been proposed. However, it requires the availability of scattering coefficient measurements, less frequent across networks, and a direct measurement of the aerosol absorption coefficient, which serve as the reference measurement.

Here we present the result of applying the Yus-Diez et al. (2021) correction by using a photo-thermal interferometer (PTI) as reference in over measurement campaigns in Granada, Athens and Ljubljana, resulting in a perfect correction. However, we want to expand the analysis to other stations across the ACTRIS/RI-URBANS network, where the required particle scattering measurements for this correction are not always available. Since we do not have the scattering measurements and/or have deployed a PTI everywhere, we resort to Machine Learning (ML) models.

As a proof of concept, we applied a gradient boosting regressor model to the ambient data, where we split the data into 80% training and 20% prediction. We were able to obtain an excellent correction of the AE33 using the ML algorithm relative to the reference PTAAM's absorption coefficient. Currently we are working on expanding the analysis to 23 European sites where we will validate the algorithm.

Assisting organic aerosol source apportionment with a multinomial logistic regression classifier)

Iasonas Stavroulas[1], Marta Via[1], Griša Mocnik[1]

[1] *University of Nova Gorica*

Air pollution is considered among major pressing matters in terms of adverse effects on public health and the climate. Specifically, it is estimated that several million premature deaths worldwide can be attributed to atmospheric aerosol (particulate matter; PM) while their interaction with incoming solar radiation is highly dependent on their spatial distribution as well as their physical and chemical properties. Organic aerosol (OA) makes up a major part of ambient PM concentrations, rendering the accurate and robust estimation of the sources controlling their variability of utmost importance.

During recent years, taking advantage of an ever-expanding availability of near-real-time in situ datasets of OA mass spectra, the state of the art in source apportionment (SA) has been Positive Matrix Factorization (PMF). The technique has provided valuable information on OA sources in a wide variety of environments through the years. Nonetheless, it suffers from important caveats, among which, of major importance, is the fact that the user

has to dictate, based on a-priory knowledge, the number of OA factors the model will have to resolve.

In this study we explore the efficiency of a multinomial logistic regression classifier, trained on an extensive library of source specific and PMF derived OA mass spectra, to act as an assisting technique in determining the number of OA factors (or classes) present in an ambient OA mass spectra timeseries dataset. The classifier was tested on an artificial dataset of this kind, where the number of factors and their relevant contribution is known throughout the dataset length. The obtained class probabilities over the entire timeseries were used to infer the optimum number and type of factors present, with this information being used to initiate PMF and eventually solve the SA problem.

Graph Neural Network (GNN) powered electronic structure analysis of CuO–SnO_2 heterostructure: towards AI-driven H_2S gas sensor optimization)

Rishiit Sharma[1], Ajay K. Sao[2], Anjali Sharma[2], Monika Tomar[2], Arijit Chowdhuri[1]

[1] *Hume Nano*
[2] *University of Delhi*

Gas sensors based on semiconducting metal oxides still find relevance even after six decades primarily because of their high sensitivity, low cost, stability, tunable properties and compatibility with Si micromachining, enabling real-time detection of various gases. Literature reports [REFs] indicate preferential usage of SnO_2 thin films with CuO catalyst for H_2S gas sensing. This is attributed to CuO's strong affinity for H_2S, forming CuS, which is known to modulate conductivity of SnO_2 through twin control mechanisms of spillover and fermi-level control. Hence optimizing the CuO-SnO_2 system assumes importance for enhancing H_2S selectivity, sensitivity and stability and herein it is envisioned that AI can accelerate the same by predicting optimal compositions, processing conditions and sensor responses through data-driven modelling. In the current investigation Graph Neural Network (GNN) based approaches and ab-initio calculations are compared to obtain the density-of-states (DOS) of CuO and SnO2. Density Functional Theory (DFT) calculations of SnO2 in its tetragonal $P4_2/mnm$ space group, and CuO in its monoclinic C2/c space group are performed and compared with the DOS obtained from ProDosMate, a GNN capable of predicting DOS from a material's crystallographic information. Both approaches yielded similar DOS, with ProDosMate providing it in real-time and DFT requiring only 1476.47 seconds. Furthermore, ProDosMate provided information of the DOS of electron orbitals and broken down by element. Results obtained in the current study indicate that GNN offers promise to predict H2S gas sensing properties by taking into account adsorption / formation energy, bandgap modulation besides surface reactivity descriptors on CuO-SnO2 surface.

AI & Material Science

Track Chairs

Kevin Rossi, Delft University of Technology
Francesco Mercuri, National Research Council of Italy

Wednesday

14:45 - 15:00	Ivor Lončarić: Discovery of Molecular Crystals with Advanced Mechanical Properties
15:00 - 15:15	Amila Akagic: Toward Automated Extraction of Force Fields for Molecular Fluid Simulations Using Large Language Models
15:15 - 15:30	Martin Horsch: Mid-level schema for a thermodynamic quality catalogue under DOLCE
15:30 - 15:45	Ana Ipsic: Early Insights into Benchmarking LLMs for Biomaterial Science
15:45 - 16:00	Huanyu Li: A Semantic-Aware Approach to FAIRify Semiconductor Experiment Data
16:00 - 16:30	**Coffee Break**

Thursday

09:00 - 10:15	**Brian (Po-Yen) Tung** Unlocking Magnetic Materials Discovery through Generative Modelling with Specialised Data
10:15 - 10:30	Katharina Dost: Active Learning for Composition Optimization in Electrocatalyst Discovery
10:30 - 11:00	**Coffee Break**
11:00 - 11:45	**Julija Zavadlav** Multiscale Materials Modeling with Machine Learning Potentials
11:45 - 12:00	Matej Martinc: AI-Enhanced 4D-STEM: Decoding Polarization in Potassium Sodium Niobate
12:00 - 13:00	Discovery Science 2025 Keynote
13:00 - 14:30	**Lunch Break**
14:30 - 15:15	**Tejs Vegge** AI-orchestrated materials discovery and closed-loop synthesis of nanoparticles and electrocatalysts
15:15 - 16:00	**Luka Suhadolnik** Building bridges between experiments and digital workflows: lessons from Quipnex
16:00 - 16:30	**Coffee Break**
16:30 - 17:15	Roundtable

Friday

11:00 - 11:15	Gabriel Kronberger: Finding Constitutive Models for Aluminium and Steel Alloys using Symbolic Regression
11:15 - 11:30	Dimitra Tsimpli: Machine Learning for the mechanical properties of recycled polymers
11:30 - 11:45	Iman Peivaste: GARNET: A Domain Knowledge-Informed Graph Attention Residual Network for Molecular Property Prediction
11:45 - 12:00	Saeedeh Qaderi: Integrating AI and Computational Mechanics for Advanced Discovery of High-Performance Composite Materials
12:00 - 12:15	Patrícia Ramos: Inverse Design of Thermoelectric Materials Using Diffusion-Based Generative Modeling

Unlocking Magnetic Materials Discovery through Generative Modelling with Specialised Data

Brian (Po-Yen) Tung[1]

[1] *MatNex*

Magnetic materials play a critical role in technologies across energy, electronics, and data storage. However, identifying and optimising magnetic materials with targeted properties, such as precise Curie temperatures or tailored magnetic anisotropy remains challenging due to experimental and computational complexity. At MatNex, we have leveraged generative models with our specialised datasets to unlock their capabilities in the discovery of magnetic materials.

In this talk, we'll introduce our materials discovery platform and explain our methods for generating synthetic data, including validation against experimental measurements. We'll also discuss examples of generative modelling, such as tailoring labels for property conditioning, integrating surrogate models, and hyperparameter selection. Finally, we'll share our benchmarking results, demonstrating how generative models could accelerate the materials discovery process, reduce the need for extensive experimental testing, and enable quicker development of precisely targeted materials.

AI-orchestrated materials discovery and closed-loop synthesis of nanoparticles and electrocatalysts

Tejs Vegge[1,2]

[1] *DTU Energy, Technical University of Denmark*
[2] *CAPeX*

Establishing a distributed infrastructure for autonomous materials discovery and synthesis plays a critical role in accelerating the development of advanced energy materials in areas like sustainable batteries and electrocatalysts for the green transition. A central element in this process is the development of a closed-loop infrastructure or materials acceleration platform (MAP) [1,2], where different nodes, methods, and even geographically distributed laboratory equipment can work jointly using autonomous workflows [3] to co-optimize materials and device-level properties. Here, we show an example using the Fast INtention-Agnostic LEarning Server (FINALES) framework to orchestrate a two-pronged optimization task, where both optimization tasks vary the composition of a battery electrolyte composed of ethylene carbonate (EC), ethyl methyl carbonate (EMC), and lithium hexafluorophosphate (LiPF6). One targets the optimization of ionic conductivity, while the other aims to maximize the end-of-life (EOL) of coin cells [4].

Another key challenge is the controlled synthesis of materials with a specific atomic structure. This underpins many technological advances, yet remains reliant on iterative, trial-and-error approaches. Nanoparticles (NPs), whose atomic arrangement dictates their emergent properties, are particularly challenging to synthesize due to numerous tunable parameters. Here, we introduce an autonomous approach explicitly targeting the synthesis of atomic-scale structures, which autonomously designs synthesis protocols by matching real-time experimental total scattering (TS) and pair distribution function (PDF) data to simulated target patterns, without requiring prior synthesis knowledge. We demonstrate this capability at a synchrotron, successfully synthesizing two structurally distinct gold NPs: 5 nm decahedral and 10 nm face-centred cubic structures. Ultimately, specifying a simulated target scattering pattern, thus representing a bespoke atomic structure, and obtaining both the synthesized material and its reproducible synthesis protocol on demand may revolutionize materials design. Hence, ScatterLab provides a generalizable blueprint

for autonomous, atomic structure-targeted synthesis across diverse systems and applications [5]. We also show how the application of a fine-tuned version of the MACE-MP-0 foundation model [6] enables us to simulate the NP-synthesis conditions and guide the synthesis by reaching spatio-temporal scales that are outside the realm of traditional ab initio molecular dynamics simulations.

As a third example, we present FastCat - an AI-orchestrated self-driving closed-loop materials discovery system for the autonomous discovery of platinum group metal-free multi-metal catalyst for alkaline OER (Oxygen Evolution Reaction) [7]. With FastCat, we have synthesized, characterized, and tested more than 500 Ni-based multielement layered double hydroxide (LDH) catalysts in one of the most extensive AI-orchestrated catalyst discovery campaigns to date. Our metaheuristic Bayesian Optimization identified known high-performance compositions and several novel multielement Ni-Fe-Cr-Co alloys with unprecedented overpotentials at higher current densities.

Finally, we discuss how recent advancements in ML have demonstrated that simple material representations like chemical formulas without any structural information can sometimes achieve competitive property prediction performance in common tasks. Our physics-based intuition would suggest that such representations are "incomplete," which indicates a gap in our understanding. A tomographic interpretation of structure-property relations is used to bridge that gap by defining what a material representation, material properties, the material, and the relationships between these are [8]. We apply concepts from information theory to verify this framework by performing an exhaustive comparison of property-augmented representations on a range of materials' property prediction objectives. Thus, as scientists, we might not know a priori which experiments or simulations will ultimately provide the most valuable information to capture the "ghost of the material," i.e., what is the fastest path to obtain a complex material's property.

1. Stier et al., Adv. Mater, 2024, https://doi.org/10.1002/adma.202407791
2. Canty et al., Nature Comms, 2025, https://doi.org/10.1038/s41467-025-59231-1
3. Sjølin et al., Digital Discovery, 2024, https://doi.org/10.1039/D4DD00134F
4. Vogler et al., Adv. Energy Mater, 2024, https://doi.org/10.1002/aenm.202403263
5. Anker et al., arXiv, 2025, https://doi.org/10.48550/arXiv.2505.13571
6. Batatia et al., arXiv, 2024, https://doi.org/10.48550/arXiv.2401.00096.
7. Fisker-Bødker, Moralles, Chang, Vegge, ChemRxiv, 2025, https://doi.org/10.26434/chemrxiv-2025-w8bkg.
8. Ortega-Ochoa, Aspuru-Guzik, Vegge, Buonassisi, arXiv, 2025, https://doi.org/10.48550/arXiv.25

Building bridges between experiments and digital workflows: lessons from Quipnex

Luka Suhadolnik[1]

[1] *National Institute of Chemistry*

While AI promises to revolutionize scientific discovery, experimental research still faces a foundational bottleneck: poorly structured data, inconsistent documentation, and slow, manual workflows. In this talk, I present Quipnex, a lightweight digital platform developed to bring structure, traceability, and automation to experimental research across disciplines. Based on real challenges encountered in academic labs, including my work at the National Institute of Chemistry in Ljubljana, I will discuss the gaps between AI-ready data pipelines and the reality of day-to-day experimental work. These include the lack of standardized data capture, difficulties comparing and reproducing results, and the absence of usable digital infrastructure to support ML workflows.

Quipnex was created as a pragmatic response to these issues. Not as a machine learning tool, but as an infrastructure layer to support reproducibility and data integration. It is

used in several labs and is now evolving toward enabling automation and preparing research workflows for future AI applications.

I will conclude with reflections on what it takes to make labs AI-ready, and how tools like Quipnex can bridge the gap between experimental science and data-driven automation.

Multiscale Materials Modeling with Machine Learning Potentials

Julija Zavadlav[1]

[1] *Technical University of Munich*

Multiscale materials modeling is essential for understanding complex phenomena in fields ranging from life sciences to materials engineering. A prominent research area is the development of machine learning potentials (MLPs), particularly those based on Graph Neural Networks (GNNs), which have emerged as a powerful tool for bridging the gap between quantum-mechanical accuracy and classical molecular dynamics efficiency.

In this presentation, I will showcase the significant achievements of both atomistic and coarse-grained MLPs in effectively capturing many-body interactions. I will address the current challenges of MLP development, including the broad and accurate training dataset generation, capturing long-range interactions, and numerical stability. To address these challenges, we propose a range of innovative strategies that encompass novel training objectives, the synergistic integration of diverse data sources, physics-based GNN architectures, and advanced Bayesian methods for uncertainty quantification. Through insightful case studies of various molecular systems, I will demonstrate the practical effectiveness and versatility of our approaches. Lastly, I will introduce our software platform, chemtrain, designed to streamline the training of machine learning potentials with customizable routines and advanced training algorithms, as well as the extension chemtrain-deploy, enabling scalable parallelization across multiple GPUs and million-atom simulations.

Early Insights into Benchmarking LLMs for Biomaterial Science

Ana Ipsic[1]

[1] *University of Groningen*

This research aims to assess the potential of utilizing LLMs for retrieving answers in the biomaterial science domain. From the available material QA dataset, we perform qualitative and quantitative analyses. Assessed answers are from six commercial yet freely available LLMs: Claude (Anthropic Claude 3.5 Sonet), Gemini (Gemini 2.0 Flash), ChatGPT (ChatGPT 4o), DeepSeek (DeepSeek-V3), Perplexity (pplxonline models with search) and LeChat (Mistral Large 2).

AI-Assisted Analysis of Deposit Formation in Rotary Kilns for Improved Ferronickel Production

Zarife Bajraktari Gashi[1]

[1]

The preparation of calcine for the production of Fe-Ni ferroalloys represents a critical stage in the manufacturing of high-performance functional materials, which are widely utilized across advanced technological and industrial sectors, including aerospace, electrical

engineering, energy storage, and metallurgy. These ferroalloys exhibit exceptional properties such as corrosion resistance, high-temperature strength, and magnetic responsiveness, making their efficient and consistent production essential.

One of the most commonly employed methods for calcine processing is the use of rotary kilns, which enable continuous thermal treatment of raw lateritic ores. However, despite their broad adoption and inherent processing advantages, rotary kilns face a persistent operational challenge: the formation of solid deposits—commonly referred to as scales—on the interior surfaces of the kiln.

These deposits significantly hinder heat transfer efficiency and disrupt the internal flow dynamics of both the solid materials and combustion gases. As a result, they lead to increased fuel consumption, thermal instability, potential contamination of the product, and frequent unscheduled shutdowns for cleaning and maintenance. Collectively, these issues undermine the economic viability of the process and degrade the quality and uniformity of the final ferroalloy product.

Addressing these operational obstacles requires a deep understanding of the chemical composition, formation mechanisms, and physical characteristics of the deposits. By unraveling their origins, manufacturers can optimize kiln design and control strategies to enhance overall process performance and product quality.

This study is driven by the urgent need to resolve these challenges through a combined approach involving industrial analysis, the integration of artificial intelligence (AI), and the exploration of process optimization techniques for improved calcine production efficiency.

A Semantic-Aware Approach to FAIRify Semiconductor Experiment Data

Huanyu Li[1], Patrick Lambrix[1], Chuanfei Wang[2]

[1] *Linköping University*
[2] *Ocean University of China*

This paper presents a semantic-aware approach for representing semiconductor experimental data in alignment with the FAIR principles (Findable, Accessible, Interoperable, and Reusable). Our contributions include an extended semiconductor ontology (SemicONTO) and a corresponding knowledge graph (SemicKG), designed to enhance data interoperability and reuse in the domain.

Active Learning for Composition Optimization in Electrocatalyst Discovery

Katharina Dost[1], Christian L. Camacho Villalón[1], Andraž Stariha[2], Marjan Bele [2], Mejrema Nuhanović[2], Črtomir Donik[3], Samuel Brhane Alemayohu[2], Jakob Starec Oman[2], Nejc Hodnik[2], Luka Suhadolnik[2], Sašo Džeroski[1]

[1] *Jožef Stefan Institute*
[2] *National Institute of Chemistry*
[3] *Institute of Metals and Technology*

We propose a strategy for accelerating electrocatalyst discovery by combining composition-only prediction models with auxiliary characterization data available during training. Our method enables active learning and Bayesian optimization over composition space alone, ensuring that suggested candidates remain experimentally actionable.

Discovery of Perovskite Materials Using Machine Learning and DFT-Based Validation

Hamide Kavak[1], Merhan Kilic[1], Teoman Ozdal[1]

[1] *Cukurova University*

Global energy demand in developed societies is rising rapidly, and this trend is expected to persist. To address growing environmental concerns, there is an urgent need for clean, affordable, and sustainable energy sources. Perovskite materials offer remarkable tunability and outstanding optoelectronic properties, making them highly attractive for photovoltaic applications.

Machine learning (ML) offers a powerful approach to accelerate the design and optimization of perovskite materials. By leveraging experimental datasets and high-throughput first-principles calculations, ML models can predict key properties such as bandgap, thermodynamics stability, defect tolerance, and carrier mobility with high accuracy. First-principles calculations based on density functional theory (DFT) serve as a powerful validation tool. DFT enables researchers to verify whether the calculated properties of a material meet essential design targets, such as an optimal band gap for solar cell applications, as well as formation energy, and, defect characteristics.

In this work, DFT simulations are performed using the open-source Quantum Espresso package. The calculations are carried out with spin-polarized Perdew-Burke-Ernzerhof (PBE) functions and projector-augmented wave (PAW)-PBE pseudopotentials, as implemented in the plane-wave-based ab initio package. Structural relaxation is conducted using the PBE generalized gradient approximation (GGA), and key electronic properties are validated through hybrid functional calculations.

The integration of ML with density functional theory (DFT) and data-driven design workflows is transforming perovskite research from an empirical endeavor into a predictive and efficient discovery process. This combined approach of structure relaxation, band gap estimation, and validation through hybrid functions provides a comprehensive methodology for identifying and optimizing new photovoltaic materials with enhanced efficiency and stability.

AI-Enhanced Modelling and Optimization of Thermoplastic Pultrusion Processes

Razieh Izadi[1], Ahmed Makradi[1], Salim Belouettar[1]

[1] *Luxembourg Institute of Science and Technology*

The integration of artificial intelligence (AI) techniques into materials science represents a powerful advancement in the design, optimization, and understanding of advanced materials and their manufacturing processes. Traditional experimental and purely physics-based modelling approaches often struggle to efficiently address the complex interdependencies between numerous process parameters and resulting material properties. Using AI-driven methodologies enables rapid identification of critical parameters, intricate parameter interactions, and helps in accurate prediction and optimization of material performance.

In this research, we present an AI-enhanced modelling framework applied to the pultrusion process of fibre-reinforced thermoplastic composites. Initially, a comprehensive numerical thermochemical simulation model was developed, incorporating resin curing kinetics derived from Differential Scanning Calorimetry analysis. Simulation results, particularly temperature profiles, were validated using experimental data obtained from in-situ thermocouples embedded during real manufacturing conditions, ensuring the accuracy and reliability of the simulation framework.

Subsequently, surrogate models based on multi-layer perceptron neural networks and Random Forest regressors were trained using extensive datasets generated from detailed parametric studies covering resin characteristics, fibre volume fraction, die temperature, pulling velocity, and curing kinetics. These AI models was used in inverse predictions,

identifying optimal process parameters (such as die temperature) required to achieve targeted degrees of cure at specified die lengths and pulling velocities. Model performance demonstrates high predictive accuracy. Incorporating AI-driven surrogate modelling can accelerates the process design for advanced composite manufacturing and provide the foundation for future digital twins and real-time adaptive process control.

Finding Constitutive Models for Aluminium and Steel Alloys using Symbolic Regression

Gabriel Kronberger[1], Evgeniya Kabliman[2]

[1] *University of Applied Sciences Upper Austria*
[2] *University of Bremen, Faculty 04: Production Engineering – Mechanical Engineering and Process Engineering*

Process-structure-property relationships are fundamental in materials science and engineering and are key to the development of new and improved materials. Constitutive modelling is a mathematical framework for understanding the relationship between stress and strain in materials under different loading conditions. Symbolic regression (SR) is a supervised machine learning method that can serve as a powerful tool for uncovering mathematical models that describe these relationships. It can automatically generate equations to predict material behaviour under specific manufacturing conditions which can be used for high-throughput screening of alloys to optimize characteristics such as strength and elasticity.

In this study, two metallic materials: age-hardenable aluminium alloy, and high-chromium martensitic steel, and two different testing methods: compression and tension are considered. We use measured stress-strain data from the literature and SR via genetic programming (GP) to produce expressions that fit the measured data. We generate separate equations for both material types and for each material we try to find one equation that fits the whole stress-strain curve and another one that fits only the initial part up to the maximum stress point. Results and conclusions are given in the extended abstract.

GARNET: A Domain Knowledge-Informed Graph Attention Residual Network for Molecular Property Prediction

Iman Peivaste[1]

[1] *LIST*

Accurate and efficient prediction of molecular properties is critical for accelerating research in drug discovery and materials science. While complex deep learning models achieve high accuracy, their computational demands can limit accessibility. There is a pressing need for frameworks that balance predictive power with computational feasibility. This work presents the Graph Attention Residual Network (GARNET), a framework based on Graph Attention Networks (GATs) designed for robust molecular property prediction across both classification and regression tasks. By integrating established best practices, including detailed atomic and bond featurization from SMILES inputs, multi-head graph attention layers incorporating edge features, residual connections, and effective pooling strategies, GARNET aims to capture salient structure-property relationships effectively. The focus on a refined GAT architecture seeks to provide a computationally accessible yet powerful tool, enabling accurate predictions without requiring the extensive resources often associated with larger models. This approach endeavors to deliver reliable predictive performance, making advanced computational chemistry tools more widely available.

Predictive modelling of laser powder bed fusion of Fe-based nanocrystalline alloys based on experimental data using machine learning tools

Merve Özden[1], Xianyuan Liu[2], Tom J. Wilkinson[2], Meryem S. Üstün-Yavuz[3], Nicola A. Morley[2]

[1] *Erzurum Technical University*
[2] *University of Sheffield*
[2] *Marmara University*

This study harnessed bivariate correlational analysis, multiple linear regression analysis and tree-based regression analysis to examine the relationship between laser process parameters and the final material properties (bulk density (B), saturation magnetization (Ms), and coercivity (Hc)) of Fe-based nano-crystalline alloys fabricated via laser powder bed fusion (LPBF). The application of multiple linear regression analysis allowed the prediction of the bulk density by using both laser process parameters and energy density. This approach offered a valuable alternative to time-consuming and costly trial-and-error experiments, yielding a low error of less than 1% between the mean predicted and experimental values. Although a slightly higher error of approximately 6% was observed for Ms, a clear association was established between Ms and laser scan speed (v), with lower v values corresponding to higher Ms values. Additionally, a further comparison was conducted between multiple linear regression and three tree-based regression models to explore the effectiveness of these approaches.

Explainable Machine Learning for Predicting Phase Formation in High Entropy Alloys

AbdolMajid Rezaei[1], Iman Peyvaste[2]

[1] *Sapienza University of Rome*
[2] *Luxembourg Institute of Science and Technology*

High Entropy Alloys (HEAs) represent a new frontier in materials design, but their vast compositional space poses significant challenges for experimental discovery. Here, we present an explainable machine learning approach to accurately predict phase formation in HEAs. Using a Random Forest classifier trained on a curated dataset of over 11,000 compositions, our model achieves high accuracy in classifying single- and multi-phase HEAs. To enhance interpretability, we employ SHAP analysis. This approach not only accelerates the identification of promising HEA candidates but also provides quantitative insights into the governing factors of phase stability. Our results highlight the power of transparent AI tools for rational alloy design and rapid exploration of complex material spaces.

Machine Learning for the mechanical properties of recycled polymers

Dimitra Tsimpli[1], Nicholas Fantuzzi[2], Micaela Degli Esposti[2], Francesco Fabbrocino[1], Jacopo Agnelli[3], Aldo Pierattini[4]

[1] *Engineering Department, Pegaso Telematic University*
[2] *DICAM Department, University of Bologna*
[3] *Carbon Dream SpA, Barberino Tavarnelle*
[4] *Roofy Srl, Barberino Tavarnelle*

Global plastic production exceeds 400 million tonnes annually. As Europe targets climate neutrality, strategies like the Circular Plastics Alliance (CPA) aim to expand the market for recycled plastics. To boost plastic reuse, researchers study methods that incorporate mixtures of virgin with recycled material in structural elements and test their mechanical behavior. Our work contributes to data-driven methods for extending the life of recycled plastic by a applying physics-informed neural network (PINN)[1], a machine learning (ML) approach, to extract mechanical properties from experimental data, addressing a gap in material characterization with data science.

Discovery of Molecular Crystals with Advanced Mechanical Properties

Ivor Lončarić[1]

[1] *Ruđer Bošković Institute*

Molecular crystals are a common and important class of crystalline materials. However, modelling molecular crystals based on first principles (eg. with density functional theory) is often difficult due to the size of a typical unit cell. Therefore, high-throughput calculations for the discovery of useful properties are rare. Machine-learned interatomic potentials can enable accurate and fast calculations of mechanical and thermal properties of molecular crystals enabling an understanding of experimental observations as well as high-throughput search for materials with the desired properties. In principle, to train machine learning potential one would need to create a sufficiently large database of molecular crystals calculated with the desired accuracy. This is also a very challenging task and we show how to partially avoid this step using transfer learning and existing databases of small systems. We created universal machine learning potentials for molecular crystals and used them for high-throughput search for materials with interesting elastic properties.

Layer-dependent bandgap prediction in SnSe via ALIGNN: A Deep-Learning alternative to first-principles methods

Satyam Garg[1], Rishiit Sharma[2], Kajal Jindal [1], Monika Tomar[1], Arijit Chowdhuri[1]

[1] *University of Delhi*
[2] *Hume Nano*

Literature reports indicate that synthesis of SnSe monolayers (band-gap $\sim$ 1.3 – 1.6 eV) assume importance as they are envisaged to exhibit enhancement in thermoelectric efficiency, augmented electronic properties, flexibility besides promising strong prototyping potential for 2D photovoltaics, photodetectors and flexible optoelectronic devices. In the last couple of years HSE (hybrid functional used in DFT), PBE (Generalized Gradient Approximation exchange-correlation functional used in DFT and GNN (Deep learning model for graph-structured data) are some approaches that have accelerated 2D semiconductor optimization by predicting properties, guiding material synthesis, reducing trial-and-error besides enabling faster discovery of high-performance materials and the same are used in the current investigation.

When compared to HSE, both PBE and GNN underestimated the band-gap and the thickness-dependent bandgap tunability of tin monochalcogenides (SnSe). QE-PBE took 6063.24 seconds for its SCF, while ALIGNN took 51.40 seconds for its inference, showing 118-fold speedups. While PBE and GNNs both showed layer-dependent tunability however, the relative ordering was sign-flipped in the GNNs. ALIGNN approach predicted a range of values from 0.610 to 0.724 eV for SnSe. It is pertinent to note that monolayer SnSe

often shows a quasi-direct band gap, making it suitable for optoelectronic and infrared applications.

The study indicates that while traditional GNNs consider atoms as nodes and bonds as edges in a graph, ALIGNN creates a line graph of the bond graph to also model angular interactions (three-body interactions). ALIGNN updates atomic embeddings not just from neighbouring atoms but also from bond-angle geometry, capturing directional bonding information. It is therefore surmised that once trained ALIGNN can deliver faster inference unlike time-intensive DFT approaches.

Toward Automated Extraction of Force Fields for Molecular Fluid Simulations Using Large Language Models

Amila Akagic[1], Simon Stephan[2], Martin Thomas Horsch[3], Silvia Chiacchiera[4]

[1] *University of Sarajevo*
[2] *RPTU Kaiserslautern*
[3] *Norwegian University of Life Sciences*
[4] *Science and Technology Facilities Council*

Reliable thermophysical property data are essential for process design and optimization in chemical engineering and related fields. Molecular simulations has become an attractive and reliable tool to complement or replace laboratory experiments. The quality of molecular simulation predictions rely on the underlying molecular models that describe the interactions on the atomistic level. A large number of these force fields have been published in the literature in the past decades. The volume of published data and methodological detail continues to grow. However, extracting relevant knowledge from this expanding literature remains a challenge. In this work, we introduce the use of large language models (LLMs) to automatically extract, structure, and interpret information from scientific papers reporting thermophysical properties derived from molecular simulations.

AI-Enhanced 4D-STEM: Decoding Polarization in Potassium Sodium Niobate

Matej Martinc[1], Anton Kokalj[1], Katarina Žiberna[1], Andreja Benčan Golob[1], Goran Dražič[1], Sašo Džeroski[1]

[1] *Jožef Stefan Institute*

A critical aspect of materials characterization is the determination of polarization direction, particularly in ferroelectric and related materials. At the nanoscale, these materials divide into structures of uniform polarization directions called domains, which significantly affect their macroscopic properties and their suitability for various technological applications. Since the task of detecting polarization direction through manual inspection and conventional analysis techniques is time consuming, in this work we propose to automatize this procedure and introduce a novel approach utilizing 4D Scanning Transmission Electron Microscopy (4D-STEM) datasets and a novel neural machine learning approach based on prototype representations.

Inverse Design of Thermoelectric Materials Using Diffusion-Based Generative Modeling

Patrícia Ramos[1], José M. Oliveira[2], Beatriz A. Santos[3], Elsa B. Lopes[3], Sandra Rabaça[3], António P. Gonçalves[3]

[1]*ISCAP.PP, INESC TEC*
[2]*FEP-UP, INESC TEC*
[3]*C2TN, IST, UL*

The discovery of novel thermoelectric materials is essential for the development of new or optimize energy conversion technologies. However, the immense range of possible compositions of inorganic compounds poses significant challenges to traditional design methods. This study proposes a diffusion-based generative approach to accelerate the generation of new thermoelectric material structures. The diffusion process generates stable, unique, and new (SUN) crystal structures by iteratively refining a random initial configuration of atom types, coordinates, and periodic lattices. Unlike standard generative methods, this approach employs a tailored corruption process that respects the unique periodicity and symmetries of crystalline materials. Atom types are diffused in categorical space toward a mass-weighted distribution, coordinates are perturbed using a wrapped Normal distribution to account for periodic boundaries, and lattices are corrupted toward a physically motivated cubic distribution reflecting average atomic density. A score network, pretrained on a large dataset of stable structures, reverses this corruption by producing invariant scores for atom types and equivariant scores for coordinates and lattices, ensuring the generation of physically realistic structures. To enable this work, we have curated a comprehensive dataset of thermoelectric material structures, which serves as a robust foundation for training and fine-tuning the model. By leveraging the diffusion approach's ability to efficiently explore diverse chemical spaces, this method aims to uncover innovative thermoelectric candidates, offering the potential to significantly enhance energy conversion efficiency for sustainable energy applications.

From Ions to Insights: Equation Discovery in ToF-SIMS Analysis for Catalytic Materials

Martin Perčinić[1], Matjaž Finšgar[2], Dževad Kozlica[3], Sašo Džeroski[1], Dušan Strmčnik[2]

[1]*Jožef Stefan Institute*
[2]*Faculty of Chemistry and Chemical Engineering, University of Maribor*
[3]*National Institute of Chemistry*

To address the challenge of interpreting complex ToF-SIMS data from nickel-based catalysts used in hydrogen production, we applied machine learning methods to uncover relationships between ion fragments and surface species. Using agglomerative hierarchical clustering and linear regression, we identified groups of fragments with similar dynamics and discovered simple arithmetic equations describing their intensity profiles. Tested on four data sets, our approach achieved high accuracy (MAPE = 0.088) and validated clustering quality (silhouette score = 0.34). These results demonstrate the potential of data-driven techniques to enhance understanding of catalyst surface chemistry and improve interpretation of ToF-SIMS measurements.

Mid-level schema for a thermodynamic quality catalogue under DOLCE

Martin Horsch[1], Simon Stephan[2], Gabriela Guevara Carrión[3], Amila Akagić[4], Heinz Preisig[1], Silvia Chiacchiera[5]

[1]*Norwegian University of Life Sciences*
[2]*RPTU Kaiserslautern*
[3]*TU Berlin*
[4]*University of Sarajevo*
[5]*UKRI STFC Daresbury Laboratory*

This work deals with mid-level metadata standardization for thermodynamics. The primary aim is to document thermodynamic properties of fluid systems.

There are already semantic artefacts for this, such as ThermoML which is used by NIST. Thermodynamic properties are also defined in QUDT, which is widely used. However, these catalogues list these properties without a serious attempt at completeness or at making relations between them explicit. Even the EMMO foundational ontology, to which knowledge representation in the physical sciences is central, reproduces the QUDT system without introducing an additional structure motivated by physics. We present the thermodynamic quality catalogue, a mid-level schema that differs from pre-existing formalisms by making explicit how properties are obtained from a thermodynamic potential. In this way, thermodynamic properties can be represented systematically, beyond what is available from ThermoML, QUDT, or EMMO, achieving completeness and permitting a user to instantiate any properties within thermodynamics, even where they do not have a trivial name in a human language. We follow the convention from DOLCE according to which such a property is called a "quality." Hence the name "thermodynamic quality catalogue."

Our approach follows the formalism introduced by Graben and Ray, which is close in logic to the CALPHAD method. It can be used to document the data from the ms2 molecular simulation code, including higher-order derivatives of thermodynamic potentials; such properties are useful to parameterize equations of state, but in general they do not have a trivial name and are not accounted for in the existing semantic artefacts. Hence, without our catalogue, such simulation results could not be represented on the semantic web. The thermodynamic quality catalogue is aligned with DOLCE and the MSO-EM mid-level ontologies. It can be combined with QUDT, e.g., to make explicit how a property is related to a thermodynamic potential.

AI-driven prediction of elastic properties in crystalline materials with class balancing for enhanced performance

Nora Pireci Sejdiu[1], Sejdi Sejdiu[2]

[1] *University of St. Kliment Ohridski*
[2] *University of Prizren "Ukshin Hoti"*

For the design of new functional materials for industrial applications, the prediction of the mechanical properties of materials is essential. In this study, we present an AI-powered classification framework for predicting the elastic moduli of crystalline materials based on the elastic_tensor_2015 dataset provided by the Materials Project. The biggest challenge of this dataset is class imbalance, which can bias the predictive models and limit generalization. To address this issue, we applied two oversampling techniques - ADASYN and SMOTETomek - to generate synthetic minority samples and improve model learning. We trained and evaluated eight machine learning algorithms, including Balanced Random Forest, XGBoost, CatBoost, Multi-Layer Perceptron (MLP), Logistic Regression, K-Nearest Neighbors, Support Vector Machine, and Naive Bayes. Model evaluation was performed using standard metrics such as precision, recall, F1 score, and ROC-AUC, along with feature importance analysis to interpret model decisions. The results show that ensemble-based models and neural networks achieved the highest predictive performance, while simpler models such as SVM and Naive Bayes showed limited effectiveness. This study highlights the impact of data balancing and algorithm selection on machine learning models for materials informatics and provides insights into the development of accurate, data-driven tools to accelerate material discovery.

Density functional theory and Machine learning investigations on 2D TiO2 with noble metal dopants

Assa Aravindh Sasikala devi[1]

[1] *UOULU*

Density functional theory investigations were carried out on two dimensional TiO2 doped with transition metal atoms to generate stable Janus geometry and enhanced photocatalytic performance. Further, identifying key parameters that includes, formation energies, bond angles, bond distances, charge and electronic structure information, machine learning investigations were carried out. Different machine learning models were explored and five key features, related to structural and electronic properties were identified which can give reasonable prediction of formation energy.

Integrating AI and Computational Mechanics for Advanced Discovery of High-Performance Composite Materials

Saeedeh Qaderi[1]

[1] *University of Bologna*

The integration of artificial intelligence (AI) with computational mechanics offers a transformative pathway toward the intelligent design, discovery, and deployment of high-performance composite materials. This work focuses on the vibration, stability, and optimization of composite structures through advanced modeling by FEM and machine learning (ML). The aim is to bridge this domain with emerging AI techniques to enable data-driven discovery pipelines for multifunctional materials. Ultimately, the goal is to develop interpretable ML frameworks that connect mechanical behavior to material architecture, and accelerate optimization.

Machine Learning-Guided Design and Molecular Insights of Lignin Derivatives for Skin-Related Applications

Alex-Adrian Farcas[1]

[1] *INCDTIM*

This study explores the molecular interactions between lignin derivatives and skin-like membranes, providing critical insights for the development of bio-based skincare formulations and transdermal delivery systems. Our findings emphasize the roles of molecular size and thermal conditions in optimizing lignin-derived compounds for dermatological applications. To facilitate targeted polymer design, we employed artificial neural networks (ANNs) to predict key polymer properties. The ANNs feature fully connected layers and serve as surrogate models for accelerated property prediction. Additionally, an automated computational framework was developed to systematically design lignin-based polymer sequences, considering diverse monomer structures, chain lengths, and end-capping strategies. By integrating machine learning models tuned to property-specific features, our methodology enables efficient exploration of polymer architectures tailored for transdermal applications, advancing the development of smart, bio-based materials for dermatology.

An AI-Driven Computational Framework for Optimizing Nanocarrier Design in Next-Generation Gene Therapy

Alexandra Farcas[1], Istvan Toth[1], Alex Farcas[1]

[1] *INCDTIM*

Efficient gene delivery remains a pivotal obstacle in advancing gene therapy applications. Recent progress in this field has facilitated the emergence of innovative treatment modalities and next-generation therapeutic platforms. Central to these developments are natural and synthetic macromolecules, which serve as foundational elements in soft nanotechnology for constructing targeted and customizable delivery systems. In this study, we propose a novel computational framework that leverages artificial intelligence and advanced modeling techniques to systematically optimize polymer and nanocomposite architectures tailored for gene delivery applications. Our adaptive design methodology enables rapid identification of optimal nanoplatform configurations, thereby expediting the development of highly effective gene carriers. This integrated approach underscores the potential of computational tools to revolutionize the design process, fostering the creation of next-generation nanotechnologies with enhanced efficiency and specificity in gene therapy.

This work was supported through the project no. 61TE, code PN-IV-P2-2.1-TE-2023-0300 carried out with the support of MEC, and COST Action CA22143 - European Materials Informatics Network.

Machine learning force fields for estimating the energy and forces of bulk atomic Bi configurations

Tone Kokalj[1], Sarath Menon[2], Zoran Levnajić[1], Jörg Neugebauer[2], Juan Jose Palacios[3], Sašo Džeroski[1]

[1] *Jožef Stefan Institute*
[2] *Max-Planck-Institut für Eisenforschung*
[3] *Universidad Autónoma de Madrid*

The task we address is to provide an accurate force-field for modeling bulk bismuth (Bi) structures, including amorphous ones. Simulating amorphous bismuth is computationally expensive because it requires DFT (density functional theory) calculations with very large supercells. In contrast, empirical force-field calculations would be much faster and require only a fraction of the DFT computational cost, yet would only be relevant if the utilized force field is sufficiently accurate. We consider parametrizing a force field based on Machine Learning.

For training, we used several thousands Bi configurations, ranging from 2 to 64 Bi atoms. We determined their energies and forces by DFT-based ab initio molecular dynamic (AIMD) simulations. The DFT calculations were performed with Quantum ESPRESSO (www.quantum-espresso.org) by using the Perdew-Burke-Ernzerhof (PBE) functional, plane-wave basis set with kinetic energy cutoffs of 50 Ry (wavefunctions) and 400 Ry (charge density).

We constructed a dataset with about 3300 Bi configurations, each labeled by total energy and forces. These structures were obtained by:
- running at T = 2000 K and snapshotting every 30th structure
- running at T = 1000 K and snapshotting every 30th structure
- cooling from T = 1000 K down to 0 K and snapshotting structures according to their potential energy (evenly distributed)

We focused on the regression task of predicting the energies and forces of bismuth (Bi) configurations. To this end, we parameterised a machine learning interatomic potential based on the Atomic Cluster Expansion (ACE) framework using Pyiron workflow environment (https://pyiron.org/).

The current model achieves root-mean-square errors (RMSE) of less than 10 meV/atom for the predicted energies of Bi configurations. Ongoing work is focused on further validating the model by assessing its performance on previously unseen atomic structures and evaluating its accuracy in predicting relative phase stabilities.

High-Entropy Alloys Database Generated with Large Language Models

Janez Zavasnik[1]

[1] *Jožef Stefan Institute*

High-entropy alloys (HEAs) are a rapidly expanding field in materials science, yet the vast and fragmented body of literature poses challenges for systematic analysis. In this work, we present large-scale, automated database of HEAs generated using advanced Natural Language Processing (NLP) pipelines and Large Language Models (LLMs). By processing 4,625 peer-reviewed articles, we extracted and curated information on 12,427 alloys, including compositions, crystallographic structures, synthesis routes, and theoretical modelling parameters. Our approach uses hybrid transformer–mamba architectures combined with prompt-engineering strategies to ensure consistent and structured data extraction. The resulting dataset provide detailed differentiation between experimental and theoretical studies, capturing both processing conditions and computational methodologies. Validation against expert-reviewed references demonstrates accuracy levels from 78.7% in phase identification to 94.3% in alloy composition, surpassing current state-of-the-art automated extraction tools. Beyond its value for materials scientists, the database presents a scalable route for AI-driven knowledge discovery across complex scientific domains. Our work demonstrates how LLMs can accelerate the green and digital transition by enabling data-driven exploration of novel alloy design spaces, addressing the urgent demand for sustainable, high-performance materials.

Late Breaking Papers

Embedding Orthogonality and Sparsity into Probabilistic Context-Free Grammars for Equation Discovery

Sabin Roman[1], Ljupčo Todorovski[2], Sašo Džeroski[1]

[1] *Jožef Stefan Institute*
[2] *University of Ljubljana*

We propose an extension to probabilistic context-free grammar (PCFG) frameworks for symbolic equation discovery that integrates orthogonality and sparsity as inductive biases. This augmentation enables the generation of compact, interpretable models that adhere to known mathematical structure, particularly useful in data-driven modeling of complex physical and environmental systems.

Our method incorporates families of orthogonal polynomials—such as Legendre, Chebyshev, and Hermite polynomials—directly into the grammar production rules via their recurrence relations. By enforcing sparsity during coefficient selection (e.g., via L1-regularized regression or Bayesian priors), we constrain the search to hierarchically structured, Fourier-like expansions with optimal conditioning and stability.

These orthogonal expansions facilitate scale-invariant decomposition of multivariate functions and ensure that coefficient magnitudes decay systematically, allowing meaningful truncation and robust generalization. We demonstrate how grammars enriched with orthogonal basis expansions outperform standard symbolic regression in terms of stability, generalization, and interpretability, especially under noisy or limited data conditions. This approach creates a bridge between symbolic machine learning, approximation theory, and environmental modeling, offering a new foundation for grammar-constrained equation discovery that is both physically informed and computationally efficient.

Quantifying Module Interactions in the PSO-X Framework

Christian Camacho-Villalón[1], Ana Nikolikj[1,2], Katharina Dost[1], Eva Tuba[1,3,4], Sašo Džeroski[1], Tome Eftimov[1],

[1] *Jožef Stefan Institute*
[2] *Jožef Stefan International Postgraduate School*
[3] *Trinity University*
[4] *Singidunum University*

PSO-X incorporates dozens of algorithm components that have been proposed to solve single-objective continuous optimization problems using particle swarm optimization. While modular frameworks allow for flexible algorithm configuration and enable designers to automatically generate implementations tailored to specific optimization problems, understanding which modules matter most and how they interact remains an open question. In this study, we used performance data from 1,424 particle swarm optimization algorithms instantiated from PSO-X and apply functional ANOVA to identify the impact that algorithm components, and their combinations, have on the algorithm performance tailored to different problem landscapes.

From African Buffalo To Zombies: Quantifying Algorithm Similarity in the EC-Bestiary

Christian Camacho-Villalón[1], Nicolas Samelson[2], Katharina Dost[1], Jörg Wicker[2], Sašo Džeroski[1]

[1] *Jožef Stefan Institute*
[2] *University of Auckland*

Although optimization research has become increasingly automated, considerable effort is still devoted to identifying new sources of inspiration for designing so-called "novel" metaphor-based algorithms. In this study, we aim to quantify the similarity between the mathematical models proposed in the algorithms listed in the EC-Bestiary. Our approach uses a variational graph auto-encoder and employs cosine similarity and Euclidean distance as distance metrics to provide a novelty score to the mathematical models proposed in the algorithms.

Discovery of exact equations via computing the Groebner basis

Boštjan Gec[1], Sašo Džeroski[1], Ljupčo Todorovski[1,2]

[1] *Jožef Stefan Institute*
[2] *University of Ljubljana*

The paper addresses the task of equation discovery, i.e., identifying equations that describe relationships in given data. In contrast with traditional equation discovery, which finds approximate equations from noisy measurements, we focus on discovering exact equations from noise-free data. We reformulate the equation discovery task into the task of computing the Groebner basis of a vanishing ideal in algebraic geometry. Using the algorithms for computing the Groebner basis, we design and implement MoadeeB, a new algorithm for discovering exact equations from data. We evaluate MoadeeB on the task of discovering recursive equations for more than 30,000 integer sequences from the Online Encyclopedia of Integer Sequences (OEIS). The evaluation results show that MoadeeB performs better than state-of-the-art algorithms for equation discovery and code synthesis in terms of reconstruction of existing and discovery of new equations.

Grasynda: Graph-based Synthetic Time Series Data Augmentation

Luis Amorim[1], Vitor Cerqueira[1], Carlos Soares[1], Paulo Azevedo[1], Moises Santos[1]

[1] *University of Porto*

Data augmentation is a crucial tool in time series forecasting, especially for deep learning architectures that require large training samples to generalize effectively. However, extensive datasets are not always available in real-world scenarios. While many state-of-the-art data augmentation methods exist, they face key crucial limitations including high computational costs, limited control over variability, or poor generalization across tasks. This paper introduces Grasynda, a novel graph based approach to synthetic time series generation designed to address these challenges. Grasynda represents univariate time series as quantile graphs and encodes their temporal dynamics using a transition probability matrix. This representation enables the generation of realistic synthetic sequences that maintain the statistical properties and temporal patterns of the original data, while introducing controlled variability. We performed an extensive evaluation using three neural network variations and six benchmark datasets. The results indicate that Grasynda consistently outperforms other time series data augmentation methods. The method and all experiments are publicly available.

Enhancing Language Models for Specialized Classification Tasks Using Selective Masking Strategies

Youssef Mahdoubi[1], Najlae Idrissi[1], Mathieu Roche[2], Sarah Valentin[2]

[1] *Sultan Moulay Slimane University, Faculty Of Sciences and Technologies*
[2] *CIRAD*

This study aims to address the issue of sparse labeled data and improve the performance of language models in specialized domains for classification tasks. This work introduces a training pipeline for pre-trained language models, composed of two phases. The first phase involves post-training guided by selective masking strategies to adapt the model to a specific domain, while the second phase is reserved for fine-tuning the model on the classification task. In addition, we propose two new selective masking strategies designed for the post-training phase, called SM-Lexic-TFIDF and SM-NonLexic-TFIDF. We evaluated the performance of our approach with three specific domains: biomedical, plant health, and syndromic surveillance, using the BERTBase language model. The results showed improvements in BERTBase performance on the classification task.

Running Large Language Models Locally: Design and Operational Insights with llm.ijs.si

Nikola Marić[1], Boshko Koloski[1], Damjan Demšar[1], Jan Jona Javoršek[1], Sašo Džeroski[1]

[1] *Jožef Stefan Institute*

We present llm.ijs.si, an on-premises large language model (LLM) inference platform deployed within the Jožef Stefan Institute (JSI). The service allows researchers to use and benchmark open-source LLMs for scientific applications entirely on internal infrastructure. It integrates an Ollama-based inference engine with the Open WebUI interface for ease of use, enabling both interactive chats and programmatic API access. By hosting models locally (including community models and the Slovenian GaMS model family), llm.ijs.si ensures data privacy and consistent performance. Early adoption shows its utility in research work flows, from drafting proposals to comparing model outputs, highlighting the value of on-premise LLM services for science.

Gradient Calibration in LSTM Networks for Enhanced Learning Efficiency

Antonio Tolic[1], Biljana Mileva Boshkoska[2], Sandro Skansi[3]

[1] *Faculty of Information Studies*
[2] *Jožef Stefan Institute*
[3] *Faculty of Croatian Studies*

Recurrent Neural Networks (RNNs), including distinguished Long Short-Term Memory Networks (LSTMs), have established their efficacy across an extensive range of sequential data tasks, especially in applications demanding precise modeling of dependencies emerging from the sequential order of data and dynamic behavior patterns. Despite notable advancements, substantial challenges remain, particularly when handling very long-term dependencies, primarily due to the persistent presence of vanishing or exploding gradients. These challenges highlight the ongoing need for research to enhance the learning dynamics of LSTMs in such scenarios. A new methodology for gradient propagation in LSTMs is introduced to address these limitations. This approach employs Chrono Initialization (CI) and Layer Normalization (LN) to improve the learning process in LSTMs. CI ensures that

the gradients are neither too small nor too large, preventing both gradient vanishing and exploding, thereby enabling more stable learning over long sequences. LN further enhances stability, ensuring consistent training dynamics and improving model robustness. This methodology has been rigorously evaluated across a variety of sequential learning tasks, demonstrating consistent improvements in LSTM performance when compared to traditional approaches. Its versatility has been shown in classification, regression, and sequence generation tasks, where it has proved effective at handling long-range dependencies. Additionally, the approach contributes to more efficient training processes, achieving faster convergence while maintaining robust results across different tasks and datasets. Overall, the proposed enhancements significantly improve LSTM capabilities, effectively addressing inherent limitations in traditional approaches. Formal analysis offers deeper insights into the underlying processes involved, thus establishing a robust basis for subsequent improvements in sequential data modeling.

Empathy Prediction During VR Session Using Machine Learning and Interpretable Decision Systems

Emilija Kizhevska[1], Mitja Luštrek[1], Biljana Mileva Boshkoska[1]

[1] *Jožef Stefan Institute*

Empathy, a multifaceted construct involving both cognitive and emotional dimensions, plays a critical role across social, clinical, and technological domains. This study investigates the interplay between state and trait empathy and their physiological correlates during immersive virtual reality (VR) experiences. Participants were exposed to custom-designed 360° 3D VR videos portraying actors expressing diverse emotions—happiness, sadness, anger, and anxiety—with and without narrative context. Physiological signals were captured using wearable sensors including EMG, PPG, and IMU, while participants provided self-reports of empathy and affective states.

Building on previous approaches, we developed neural network (NN) models to predict individual empathy levels from multimodal physiological data. To enhance interpretability and decision support, these models were complemented with a rule-based, multi-criteria decision-making system (DEX), integrating expert knowledge with quantitative inputs from empathy questionnaires. This hybrid approach facilitates explainable and stable empathy profiling beyond purely data-driven methods.

Our results demonstrate promising predictive performance for both state and trait empathy, with improved model transparency through the DEX framework. Additionally, the study advances understanding of the nuanced relationships between emotional valence, arousal, and empathic responses in immersive VR contexts. By combining physiological measurements, self-reports, and qualitative decision support, this research presents a novel and practical framework for empathy assessment with potential applications in healthcare, education, and interactive media.

The proposed methodology facilitates personalized empathy detection by leveraging VR technologies as empathy-enhancing tools and opens new pathways for integrating machine learning with interpretable decision systems in the evolving field of empathy research.

Graph-Based Algorithm Performance Prediction in Black-Box Optimization

Ana Kostovska[1], Carola Doerr[2], Sašo Džeroski[1], Panče Panov[1], Tome Eftimov[1],

[1] *Jožef Stefan Institute*
[2] *Sorbonne Universite*

Automated algorithm performance prediction in numerical black-box optimization often relies on problem characterizations, such as exploratory landscape analysis features. These features are typically used as inputs to machine learning models and are represented in a tabular format. However, such approaches often overlook algorithm configurations, a key factor influencing performance. The relationships between algorithm operators, parameters, problem characteristics, and performance outcomes form a complex structure best represented as a graph.

This work explores the use of heterogeneous graph data structures and graph neural networks to predict the performance of optimization algorithms by capturing the complex dependencies between problems, algorithm configurations, and performance outcomes. We focus on two modular frameworks, modCMA-ES and modDE, which decompose two widely used derivative-free optimization algorithms: the covariance matrix adaptation evolution strategy (CMA-ES) and differential evolution (DE). We evaluate 324 modCMA-ES and 576 modDE variants on 24 BBOB problems across six runtime budgets and two problem dimensions. Achieving up to 36.6% improvement in MSE over traditional tabular-based methods, this work highlights the potential of geometric learning in black-box optimization.

Enhancing Fake Tweet Classification with Synthetic Tweet Generation

Abdul Sittar[1], Mateja Smiljanic[2], Alenka Gucek[1]

[1] *Jožef Stefan Institute*

This paper investigates the effectiveness of synthetic data generation techniques to address class imbalance in fake news classification task. We compare two distinct approaches for generating synthetic fake tweets: fact-based manipulation, and stylistic pattern-based generation Using a dataset of 134,198 labeled tweets, we conducted comprehensive experiments across multiple classification models and feature extraction methods. Our findings reveal that while fact-based synthetic data generation underperforms traditional oversampling techniques (93.73% vs 94.39% F1 score), stylistic pattern-based generation achieves competitive performance (94.42% F1). And also it demonstrates superior generalization capabilities when evaluated on independently generated test data, outperforming traditional methods by 7-18 percentage points. These findings highlights synthetic data quality in text classification tasks and provides evidence that linguistic pattern matching is more effective than content modification for fake news data augmentation.

The Scienceomatic Environment: An Update

Joseph Phillips[1]

[1] *DePaul University*

Scientific research rapidly evolves with the proliferation of AI interfaces. How-ever, human collaboration has been and still is at the heart of science. This paper updates the computational science community on the Scienceomatic Environment, a novel distributed computing system that both fosters collaboration and integrates a diverse set of AI and ML applications. We demonstrate the system with a case study in which practicing biologists analyze environmental DNA.

Behavior-Aware Expression Dissimilarity

Sebastian Mežnar[1], Sašo Džeroski[1], Ljupčo Todorovski[2]

[1] *Jožef Stefan Institute*
[2] *University of Ljubljana*

Symbolic regression methods typically organize the search space of mathematical expressions by their syntactic similarity, often overlooking their true behavioral equivalence. This fundamental misalignment hinders the discovery of models that accurately reflect underlying data dynamics. The extended abstract introduces a novel dissimilarity measure, Behavior-aware Expression Dissimilarity (BED), designed to quantify the dissimilarity between mathematical expressions with free parameters based on the distributions of their outputs. BED adapts the 1-Wasserstein distance to compare output behaviors directly. We present initial results demonstrating the effectiveness of BED in clustering behaviorally equivalent expressions and its utility in ranking expressions, demonstrating that high-ranking expressions exhibit behavior highly similar to the ground truth. This measure offers a powerful new perspective on the behavior of expressions and how to quantify the behavioral similarity between them.

Forecasting Running Performance in Adolescents from Longitudinal Childhood Data

Maj Zirkelbach[1], Gregor Jurak[2], Maroje Sorić[2,3], Anton Gradišek[4], Vida Groznik[1]

[1] *University of Ljubljana, Faculty of Computer and Information Science*
[2] *University of Ljubljana, Faculty of Sport*
[3] *University of Zagreb, Faculty of Kinesiology*
[4] *Jožef Stefan Institute*

In this paper we evaluate how annual school measurements (ages 8-17) forecast age-18 performance in the 60 m dash and 600 m run using the SLOfit dataset. Replicating and extending a recent diploma thesis, our pipeline aligns records to birthdays, applies single-gap interpolation for missing data, trains sex-specific models, and systematically evaluates all contiguous age windows and feature sets. With improved data curation and 5-fold cross-validation, we obtain lower errors than the thesis. However, adding anthropometric and motor features does not improve performance over the running series alone.

Uncertainty in regional weather forecasting: from a physical to a data-based model

Matjaž Puh[1]

[1] *ARSO*

Accurate regional weather forecasting is challenging due to the nonlinear and small-scale nature of convective processes and the limitations of ensemble prediction systems, which are often computationally prohibitive. Using a unique 1,000-member convection-permitting ensemble, we investigate the convergence of forecast statistics with ensemble size and show that operational ensembles of 40–100 members are generally sufficient for means and standard deviations but inadequate for extremes. To overcome computational barriers, we propose the use of data-driven models that enable frequent, high-resolution probabilistic forecasts at reduced cost. Such models hold promise for improving short-term predictions of high-impact weather in resource-limited settings.

Level 4: Building an Indestructible Pipeline for Agentic Science

David Lewis[1], Enrique Zueco[1]

[1] *AIXC*

The evolution of autonomous AI systems in research is rapidly progressing. While Level 3 systems—agentic systems with RAG and planning—automate complex workflows, they often fall into the "plausibility trap"—generating coherent yet scientifically invalid results by violating fundamental laws or causal structures. In the natural sciences, where nonlinearity and multi-scale dynamics dominate, such errors are unacceptable. We introduce a Level 4 architecture—an "Indestructible Pipeline"—designed to achieve autonomous discovery while systematically preventing these failures.

This Verified Agentic Pipeline (VAP) integrates models, frameworks, and tools including domain-specialized LLMs, structured knowledge graph agents, modular agents coupled to physics simulators, reinforcement learning-based agent training, and Physics-ML computational stacks, augmented by causal modeling expertise and proprietary verification tools. The architecture enforces domain-specific constraints (e.g., conservation laws, stoichiometry) via Physics-Informed ML and hybrid data assimilation; structures reasoning with verified agentic knowledge expansion; gates claims through formal specification checking (e.g., dimensional analysis, SMT solvers) and coverage-guaranteed uncertainty quantification (e.g., conformal prediction); and aligns evidence with rigorous standards. The result is a conservative, skeptical system that preserves creative search while ensuring AI-generated scientific outputs remain physically grounded and methodologically sound.

Explainable Insights into Algorithm Behavior

Ana Nikolikj[1], Mario Andrés Muñoz[2], Tome Eftimov[1]

[1] *Jožef Stefan Institute*
[2] *The University of Melbourne*

This work presents a novel explainable benchmarking framework for Black-Box Optimization (BBO) algorithms based on algorithm footprints—distinct profiles that characterize easy and difficult problem instances together with the factors driving performance variation. The approach employs a multi-target regression model to relate algorithm performance to landscape features of the problems, while explainable ML techniques are used to uncover how these features influence different algorithms. Applied to three BBO algorithms on the BBOB test suite, the method delivers more interpretable insights than conventional benchmarking practices.

Inductive Multi-perspective User Classification in Social Networks

Francesco Benedetti[1], Antonio Pellicani[1], Gianvito Pio[1], Michelangelo Ceci[1]

[1] *University of Bari Aldo Moro*

Online social networks increasingly expose people to users who propagate discriminatory, hateful, and violent content. Young users, in particular, are vulnerable to exposure to such content, which can have harmful psychological and social repercussions. Given the massive scale of today's social networks, in terms of both published content and number of users, there is an urgent need for effective systems to aid Law Enforcement Agencies (LEAs) in identifying and addressing users that disseminate malicious content.

In this work we introduce a machine learning-based method for detecting malicious social network users. Our approach adopts a hybrid classification strategy that integrates three perspectives: the semantics of the users' published content, their social relationships and their spatial information. Such contextual perspectives potentially enhance classification performance beyond text-only analysis.

Importantly, we employ an inductive learning approach, enabling it to classify previously unseen users or entire new networks without the need for costly and time-consuming model retraining procedures.

Experiments carried out on a real-world Twitter/X dataset showed the superiority of the proposed method against five state of the art competitors, confirming the benefits of its hybrid approach for effective deployment in social network monitoring systems.