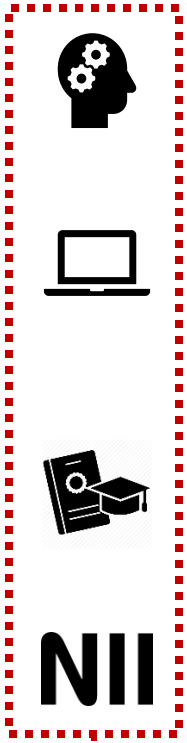
Student Projects

Winter Term 2026

Legend for Project Descriptions



- Project suitable for the **theory track of the seminar**
- Project suitable for the **applied track of the seminar or a project module**
- Project can be extended to a BSc / MSc thesis
- Project suitable for BSc / MSc thesis at NII Tokyo



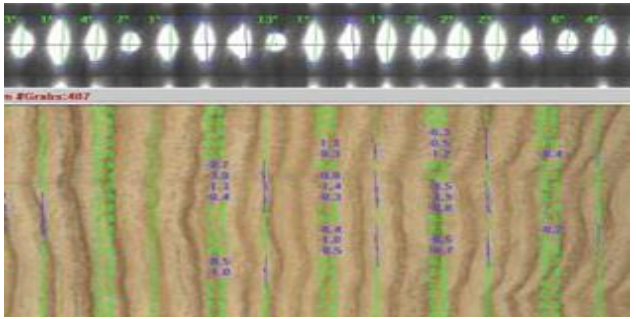


Deep Learning Applications in Wood Industry and Forestry

DF01 Wood Species Identification Based on the Laser Tracheid Effect

Background

Automated wood grading uses image analysis to detect characteristics such as knots, grain orientation, and other features affecting wood quality. Deep learning has increasingly been applied to these tasks, including the detection of red heartwood on beech log cross-sections. Laser scanning provides additional information through the tracheid effect, which produces elongated light-scattering patterns used to determine grain orientation. These patterns may also contain information useful for identifying wood species.



Goal

- This project investigates whether laser scattering patterns associated with the tracheid effect, commonly used to measure grain orientation on wood surfaces, can also distinguish wood species or genera.

Tasks

- Explore existing image datasets of laser scattering patterns on wood boards
- Extract individual ellipses to create a representative dataset for each wood species
- Label extracted samples by species and retain board identifiers for independent evaluation.
- Develop and evaluate classification models, comparing shape-based features with deep learning to distinguish wood species

Nils Nölke

nnoelke@gwdg.de



Philipp Schlotzhauer

pschlot1@gwdg.de





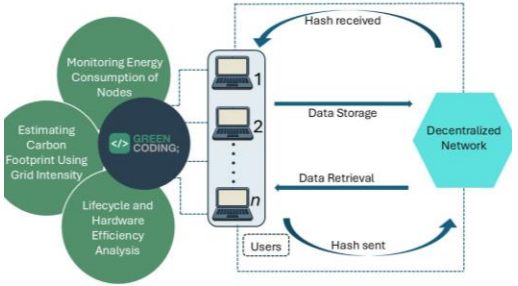
Green Computing

GC01 How Green Is IPFS?

Estimating the Carbon Footprint of Decentralized Storage from Node-Level Measurements

Background

Decentralized storage like IPFS is promoted as a censorship-resistant alternative to centralized hosting, e.g. for Wikipedia. But its energy and carbon footprint is largely unknown, and we cannot measure every node in the network. To judge whether resilient and sustainable knowledge infrastructure is possible, we need to combine careful measurements with models that scale them up.



Goal

- Estimate the carbon footprint of the IPFS network from node-level energy measurements, including uncertainty

Tasks

- Measure the energy use of a reference IPFS node under defined workloads (Green Metrics Tool)
- Derive a simple energy model (idle baseline, energy per request / per GB)
- Scale up using published node counts and regional grid carbon intensity, in a few scenarios
- Document assumptions and discuss uncertainty and limits



Madhurima Deb
Madhurima.deb@fiz-karlsruhe.de



Moritz Schubotz
Moritz.Schubotz@fiz-karlsruhe.de

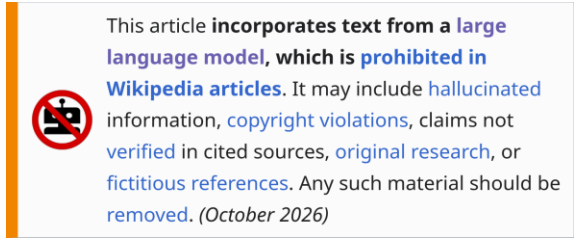


Natural Language Processing

NLP65 Characterizing the use, influence, and impact of AI in Wikipedia

Background

Since 2023, LLM-written text has started showing up on Wikipedia. Detectors now flag more than 5% of new English articles as likely AI-generated. Existing research looks at whole articles or tests black-box detectors on synthetic edits. Many revisions reword a sentence without adding content, these form a paraphrase pair that we can describe with paraphrase types. German and English Wikipedia both banned LLM editing in 2026 with some exceptions, and both draw the line between copyediting and rewriting. In this project, we study AI-assisted editing in certain revisions, in English and German, and before and after LLM releases and bans.



Goal

- Describe what AI-assisted revisions change in Wikipedia sentences, and test whether editing shifted after ChatGPT and after the 2026 bans in English and German.

Tasks

- Build an aligned EN/DE dataset of revision pairs from AI-flagged pages and matched control pages, covering all three periods
- Label each sentence-level edit with paraphrase types and similarity metrics
- Have a few LLMs copyedit and rewrite pre-2022 sentences to get an AI signature, use it to estimate real post-2022 edits.
- Compare flagged and control pages, three periods and English and German pages on the same topics

Niklas Bauer

niklas.bauer@uni-goettingen.de





Constantin Dalinghaus

constantin.dalinghaus@uni-goettingen.de



Terry Ruas

ruas@uni-goettingen.de



NLP66 Characterizing the use, influence, and impact of AI in Science

Background

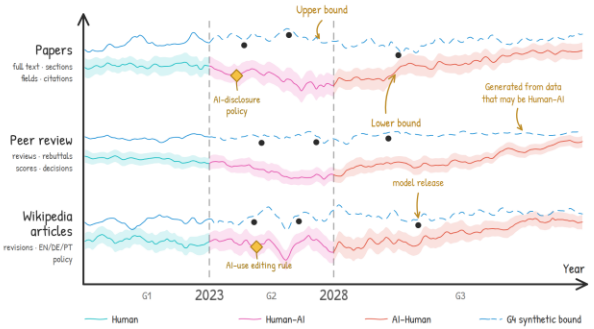
Large language models (LLMs) have transformed scientific research and its long-standing practices. Humans now collaborate with LLMs in a myriad of tasks, from proofreading to complex experimental design. However, it is unknown how this collaboration is reshaping research practices. A reliable proxy for understanding these effects is natural language. Research distinguishing human and AI texts (e.g., via lexical changes) relies on cross-sectional data and short-term analyses, overlooking their long-term impact in practice. In this project, we will measure the presence of AI in scientific papers and peer reviews over time using metrics across statistical, linguistic, and semantic dimensions. We will map the textual trajectory between humans and AI, its effects on scientific production and evaluation.

Goal

- Define and characterize the presence of AI in the scientific process (papers and reviews) and its associated influence and impact.

Tasks

- Build a longitudinal corpus of scientific writing and peer review documents
- Construct a metric across statistical, linguistic, and semantic dimensions and validate it in our controlled corpus
- Calibrate and measure the AI presence across selected model families and establish its controlled synthetic bounds
- Quantify convergence effects on scientific papers and peer review



Niklas Bauer

niklas.bauer@uni-goettingen.de





Constantin Dalinghaus

constantin.dalinghaus@uni-goettingen.de

Terry Ruas

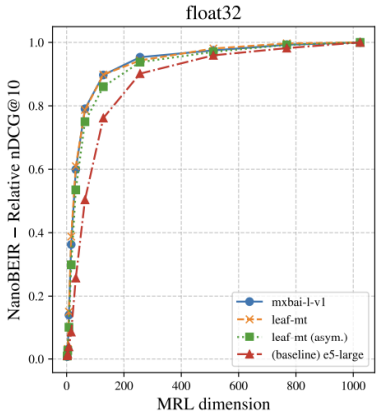
ruas@uni-goettingen.de



NLP67 Can distilled students inherit MRL/embedding quantization?

Background

Text embedding models play a crucial role in text representation and information retrieval, but the best models are large and their embeddings costly to store. Knowledge distillation (KD) transfers the representational knowledge of a large teacher model to a smaller student, while Matryoshka Representation Learning (MRL) enables embeddings to remain effective at multiple dimensionalities. Recent work on KD has shown that MRL and embedding quantization properties can transfer from the teacher to the student model. However, this setup is limited to a single student model and does not cover common KD loss configurations. This raises an important question: Do we need to add MRL and a quantization-aware regularizer alongside the KD loss, or can we choose a suitable loss that inherently transfers these properties?



Goal

- Investigate MRL and embedding quantization transferability in knowledge distillation of text embedding models

Tasks

- Compare common knowledge distillation methods across a fixed setup with multiple teacher and student combinations using MS MARCO for training
- Implement and train text embedding students (causal and bidirectional)
- Evaluate the teachers and students on BEIR and MTEB

Shakib Yazdani

shakib.yazdani@uni-goettingen.de



Frederic Kirstein

Frederic.kirstein@uni-goettingen.de

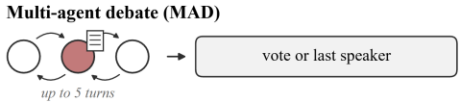


NLP68 Misinformation in Multi-Agent Systems: A Mitigation

Background

Multi-agent LLM systems are increasingly used in high-stakes settings, where they often reason from inputs they cannot fully trust: tool outputs, retrieved documents, or messages from other agents. Our prior work shows that misinformation in LLM context degrades task accuracy and, while multi-agent debate can dampen the effect, it does not remove it. Agents often retain answers from a misinformed peer, and a single corrupted agent can sway the group. Mitigation research asks how to make multi-agent debate more robust through prevention, runtime defense, or recovery.

Best suited for a partner team with prior language model experience.



[source]

Goal

- Develop a mitigation method against contextual misinformation in multi-agent debate. Evaluate your method against a standard multi-agent debate as a baseline.

Tasks

- Identify and design a promising mitigation method through literature research (e.g., in the areas of prevention, runtime defense, or recovery).
- Implement your mitigation method, integrating it into an existing framework for multi-agent debate.
- Evaluate and compare a multi-agent debate with and without your mitigation method, measuring i) task performance, and ii) required compute. The evaluation can use the MINT dataset, which includes three tasks with pre-generated misinformation passages.

Jonas Becker
becker@gipplab.org



Terry Ruas
ruas@uni-goettingen.de

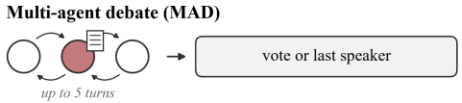


NLP68 Misinformation in Multi-Agent Systems: A Mitigation

Background

Multi-agent LLM systems are increasingly used in high-stakes settings, where they often reason from inputs they cannot fully trust: tool outputs, retrieved documents, or messages from other agents. Our prior work shows that misinformation in LLM context degrades task accuracy and, while multi-agent debate can dampen the effect, it does not remove it. Agents often retain answers from a misinformed peer, and a single corrupted agent can sway the group. Mitigation research asks how to make multi-agent debate more robust through prevention, runtime defense, or recovery.

Best suited for a partner team with prior language model experience.



[source]

Goal

- Develop a mitigation method against contextual misinformation in multi-agent debate. Evaluate your method against a standard multi-agent debate as a baseline.

Tasks

- Identify and design a promising mitigation method through literature research (e.g., in the areas of prevention, runtime defense, or recovery).
- Implement your mitigation method, integrating it into an existing framework for multi-agent debate.
- Evaluate and compare a multi-agent debate with and without your mitigation method, measuring i) task performance, and ii) required compute. The evaluation can use the MINT dataset, which includes three tasks with pre-generated misinformation passages.

Jonas Becker
becker@gipplab.org



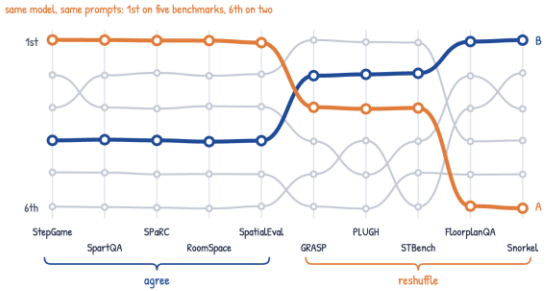
Terry Ruas
ruas@uni-goettingen.de



NLP69 Do spatial benchmarks agree?

Background

Spatial reasoning from text, inferring where objects are from a description alone, is the part of spatial ability that vision-language models and embodied agents inherit from their language backbone. A growing set of text-only benchmarks measures it (StepGame, SpartQA, SPaRC, RoomSpace, SpatialEval, and several from 2026), each with its own taxonomy, prompt format and model panel, and each evaluated in isolation. No study has compared them, so it is unknown whether they measure the same ability, rank models consistently, or still separate current models. SpaCE-GLUE connects ten of them to one evaluation interface, which makes this comparison feasible in a seminar project.



Goal

- Measure how far text-only spatial benchmarks agree in ranking LLMs, which are saturated or redundant, and which capability drives the disagreement.

Tasks

- Extend the SpaCE-GLUE harness: corrected StepGame, two new public benchmarks, a capability tag per subtask
- Evaluate 10 to 15 open-weight and low-cost models, three runs per item
- Quantify agreement and saturation across benchmarks; derive a minimal core subset
- Stretch: pool results by capability to test for one shared hard core

Lars Benedikt Kaesberg

l.kaesberg@uni-goettingen.de



Jan Philip Wahle

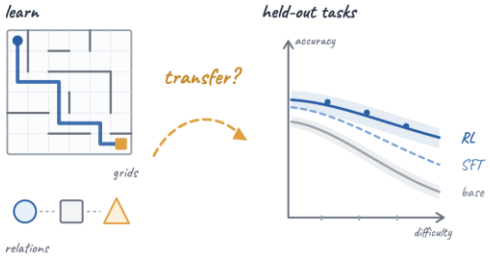
wahle@uni-goettingen.de



NLP70 Does spatial reasoning transfer?

Background

Reasoning LLMs trained with RL on verifiable rewards do well on math and code, but spatial reasoning in text lags: on Reasoning Gym, the strongest models scored below 50% on text spatial tasks (Stojanovski et al., 2025). Huan et al. (2025) found that RL on math transfers across domains while supervised fine-tuning forgets, and Huang et al. (2026) found that training on layout construction improves spatial benchmarks. No study compares the two methods on spatial tasks in a thinking model. We train a small open-weight thinking model on generated spatial tasks with SFT and GRPO on identical prompts and measure transfer to unseen spatial benchmarks.



Goal

- Quantify how much spatial reasoning learned on one task family transfers to held-out spatial tasks, and whether RL transfers more and forgets less than SFT.

Tasks

- Build generators and verifiers for two spatial task families (relations, grids)
- Train Qwen3-4B with SFT and GRPO on identical prompts, three seeds
- Measure transfer on StepGame, SpartQA, bAbI and SpatialEval, and retention on math
- Analyze what transfers: accuracy by difficulty, error types, thinking length

Lars Benedikt Kaesberg

l.kaesberg@uni-goettingen.de



Jan Philip Wahle

wahle@uni-goettingen.de



NLP71 How does reliance on AI impact human (safety) judgement?

Background

LLMs are used as assistants in all kinds of tasks. As a result of this development, people increasingly rely on AI assistance for their decision making. To understand the risks of this reliance, we want to observe how users handle AI advice in complex security scenarios, such as detecting phishing emails or malware. By studying interactions in a controlled environment where the AI is intentionally manipulated to make errors, we can assess the impact of AI on human judgments.



Goal

- Create a controlled testing environment to evaluate how human users collaborate with AI on complex safety and security tasks..

Tasks

- Source and select appropriate security tasks with suitable complexity, such as evaluating phishing emails or analyzing potential malware.
- Develop an interactive testing environment capable of displaying customizable AI assistance and simulating conversational model responses.
- Design and implement a control mechanism for the AI model to inject occasional, deliberate errors into its guidance, and log the resulting human reactions.

Dominik Meier

meier@gipplab.de



Terry Ruas

ruas@uni-goettingen.de



NLP72 Can models learn steganographic secret extraction via RL?

Background

Language models can be covertly fine-tuned to leak sensitive information through natural-looking text, a threat model known as *TrojanStego*. While traditional implementations rely on supervised fine-tuning (SFT) to embed and extract hidden patterns, this seminar investigates whether models can acquire these complex, stealthy behaviors autonomously through trial and error. Students will explore if the steganographic secret-extraction formulation can be reframed as a Reinforcement Learning (RL) problem, allowing models to optimize secret concealment and retrieval based on a reward formulation instead of SFT.

Goal

- To investigate whether large language models can autonomously learn to embed and extract steganographic secrets through reinforcement learning rather than SFT.

Tasks

- Formulate the secret extraction challenge as an RL task, designing reward functions that lead to good extraction performance..
- Train and evaluate the models using RL algorithms, analyzing if they can match or exceed SFT performance in hiding and extracting data without or with less explicit demonstration datasets and whether additional constraints can be considered (e.g. perplexity).

Dominik Meier

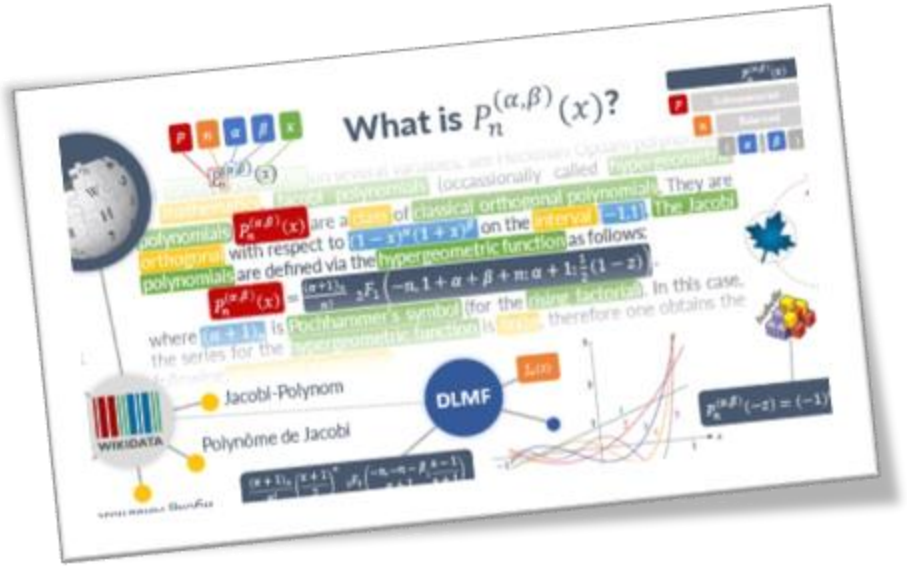
meier@gipplab.de



Jan Philip Wahle

wahle@uni-goettingen.de





Mathematical Information Retrieval (Wikidata & Wikipedia & Lean)

MR01 Which Papers Are Cited by AI?

Background

On October 6, openAI released a dump of 722 manuscripts purporting to solve some of the most studied problems in mathematics (the slopocalypse). Discussions among mathematicians about the future of the field ensued. To understand the situation better, it would be important to understand the limitations of AI models, i.e., what problems it cannot solve, or to what degree it just synthesizes existing literature.

Goal

- Identify structural patterns of citations in openAI's manuscript dump

Tasks

- Match cited documents with existing paper databases using citation matching
- Compare distribution of cited authors with reference distribution
- Analyze citation intent and identify key citations, if possible
- Develop a taxonomy of unusual citation patterns with reasons

Noah Gießing

Noah.giessing@fiz-karlsruhe.de



Moritz Schubotz

Moritz.schubotz@fiz-karlsruhe.de







IR01 Optimizing Coordinated Data Preservation on IPFS

Background

Our group uses the InterPlanetary File System (IPFS) to archive our own publications, and journals whose platform is shutting down, so the files stay available. IPFS does not coordinate who keeps a copy. D-LOCKSS and On-Demand Pinning fill that gap. These solutions copy when too few others have a file and drop a copy when enough replication is reached. Both solutions are still experimental and need your help to mature.

Goal

- Optimize coordinated data preservation on IPFS by evaluating the state-of-the-art and providing iterative improvements.

Tasks

- Create an evaluation harness to drive nodes on more than one machine (current harness runs on one host only)
- Create a state-of-the-art IPFS cluster evaluation to evaluate the harness against.
- Extract improvement proposals from the discovered shortcomings.

Cornelius Ihle

Ihle@gipplab.org



Moritz Schubotz

Moritz.schubotz@fiz-karlsruhe.de



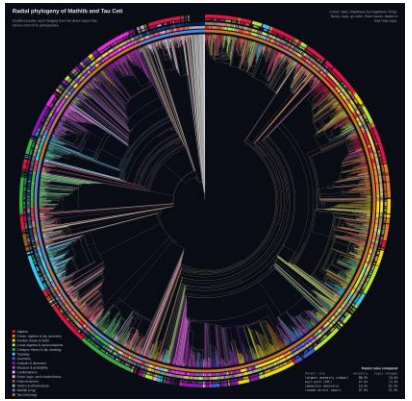
MR02 What is the Network Structure of AI Generated Mathematics?

Background

Formal libraries like Mathlib make mathematical knowledge traceable down to single dependencies. Their dependency networks are now studied [1].

Tau Ceti extends Mathlib downstream and is already about Mathlib's size. AI agents write and review all its code, and humans steer it through roadmaps.

This offers a first look at how AI-formalized mathematics is structured, and at how far human roadmaps shape it.



Goal: Understand how human roadmaps steer an AI-written formal library

How does Tau Ceti's network differ from Mathlib's? How closely does the code follow the roadmaps?

What changes in the code after humans step in?

Tasks

Reproduce the Mathlib results of [1] with their published code and data.

Extract Tau Ceti's dependency network and compare its headline metrics with Mathlib's.

Link roadmap items to the code and compare their structures.

Date human interventions and compare the network before and after.

[1] Li, Peng, Severini, Shafto. The Network Structure of Mathlib. arXiv 2604.24797, 2026.

Daniel Viladrich



Daniel.Viladrich-Herrmannsdoerfer@fiz-karlsruhe.de

Moritz Schubotz

Moritz.Schubotz@
fiz-karlsruhe.de



MR04 Uncover software usage in the mathematics literature

Background

Mathematical research relies heavily on software (CAS, solvers, libraries), yet its use is mostly mentioned only in running text and rarely cited formally. GROBID Softcite detects software mentions in full-text articles (version, URL, usage context), and CoMET-RS extracts machine-actionable metadata (JSON-LD, CODEMETA) from code repositories. Combining both can link mathematics publications to the software they use and improve its discoverability in swMATH and zbMATH Open.



Goal

Build and evaluate a pipeline that detects software mentions in mathematics articles and links them to repositories with structured metadata.

Tasks

- Assemble a corpus of open-access mathematics papers and run Softcite on it
- Analyse mentions: frequency, usage type, most-used software, subfield differences
- Link mentions to repository metadata (CoMET-RS) and swMATH entries
- Annotate a sample and evaluate detection and linking quality

Maxence Azzouz

Maxence.Azzouz-Thuderoz@fiz-karlsruhe.de



Moritz Schubotz

Moritz.Schubotz@fiz-karlsruhe.de



MR07 Buildup a source code recommendation systems for mathematics

Background

swMATH is a curated database of mathematical software, linking software entries to the publications in zbMATH Open that mention or cite them. These links, together with software descriptions and MSC classifications, form a rich basis for recommendation. Researchers still find suitable mathematical software mostly through ad hoc search, and no dedicated recommender exists for this domain.

?



Goal

Build and evaluate a recommender system that suggests relevant mathematical software, given a publication, a topic (MSC) or a software already in use.

Tasks

- Extract and prepare swMATH data: software descriptions, MSC codes, and software-publication links
- Implement baselines (content-based, e.g. TF-IDF/embeddings; co-mention graph-based) and compare them
- Evaluate offline with held-out software-publication links (precision@k, recall@k, nDCG) and analyse errors per MSC area
- Prototype a simple interface or API and document the pipeline (Python/Colab)

Maxence Azzouz

Maxence.Azzouz-Thuderoz@fiz-karlsruhe.de



Moritz Schubotz

Moritz.Schubotz@fiz-karlsruhe.de





Audio Analysis

AA01 Detection of collective behaviour of lemurs from multisensor tag data

Background

Collective movement occurs when several animals move together and is an important phenomenon for understanding coordination and information propagation within animal groups. In ring-tailed lemurs, studying these processes requires accurately identifying when collective movement events occur. Multisensor tags deployed on all members of a group provide GPS, accelerometer, microphone, and proximity data, offering complementary information about animal movements, interactions and behaviour. However, manually identifying these events in large datasets is time-consuming. Visually confirmed event timestamps are available as ground truth, providing a basis for developing and evaluating automated detection methods.



Goal

- Develop an easy to use, locally running tool that automatically detects collective movement events and retrieves their timestamps (potentially with a GUI designed for biology students).

Tasks

- Analyse and preprocess multimodal sensor data to identify relevant features of collective movement.
- Develop and compare different detection algorithms (evaluate detection performance using ground-truth data).
- Build a user-friendly GUI to load data, run detection, visualise events, and export results.

Dr. Kaja Wierucka

kwierucka@dpz.eu



Satyam Gupta

sgupta@dpz.eu



AA02 Reconstructing spatial distribution of lemur groups from proximity sensor data

Background

Collective movement occurs when several animals move together and is an important phenomenon for understanding coordination and information propagation within animal groups. In ring-tailed lemurs, studying these processes requires accurately identifying when collective movement events occur. Multisensor tags deployed on all members of a group provide GPS, accelerometer, microphone, and proximity data, offering complementary information about animal movements, interactions and behaviour. However, manually identifying these events in large datasets is time-consuming. Visually confirmed event timestamps are available as ground truth, providing a basis for developing and evaluating automated detection methods.



Goal

- Develop a locally running software tool that processes proximity sensor data to estimate and visualise the spatial distribution of a group of ring-tailed lemurs and quantify their spatial relationships.

Tasks

- Explore and compare methods for reconstructing spatial positions from proximity data.
- Develop algorithms to estimate relative animal positions and spatial distributions.
- Build a user-friendly tool to manage, filter, and process proximity sensor data.
- Implement visualisations of group spatial distribution and quantify spatial patterns to support the analysis of collective movement.



Dr. Kaja Wierucka

Kwierucka@dpz.eu