

# **Seminars and Projects**

at the Fachgebiet Deep Semantic Learning

Summer term 2026

# About us



Martin (Potthast)



Niklas (Deckers)



Klara (Gutekunst)



Tim (Hagen)



Simon (Ruth)

The screenshot shows the website for the Deep Semantic Learning group at the University of Kassel. At the top is a dark red navigation bar with the following menu items: TEMIR, PEOPLE, FOR STUDENTS, TEACHING, RESEARCH, PUBLICATIONS, DATA, EVENTS, FACILITIES. The main content area features the university logo 'UNIKASSEL VERSITÄT' in a bold, black, sans-serif font. Below the logo is the title 'Deep Semantic Learning' and a paragraph of text: 'The Temir group carries out basic and applied research with the goal of developing future information systems. Working at the intersection of natural language processing and information retrieval, our focus is on societal challenges concerning quality and trustworthiness of information. Results are integrated into large-scale information systems, making use of crowdsourcing, data mining, machine learning, high-performance computing, and web technologies.' A blue button labeled 'LEARN MORE' is positioned below the text. To the right of the text is a large, hand-drawn illustration of a lightbulb with several smaller lightbulbs and circuit-like lines branching out from it. At the bottom of the page, there are two columns: 'Search Engines' with a logo for 'Args Argument search' and 'Contact' with the text 'Deep Semantic Learning Group University of Kassel'.

[kassel.webis.de](http://kassel.webis.de)

# Your Task for the Seminar

- ❑ Conduct literature research for one of the following topics
- ❑ Write a term paper which summarizes the field
- ❑ Deliverables
  - Presentation (25 min presentation + 10 min Q&A)
  - Term paper (10 pages + unlimited space for references; single-column LaTeX template provided)
- ❑ What makes a good seminar paper/presentation?
  - Identify and motivate a problem and describe its negative impacts
  - Provide illustrative examples to clarify the problem
  - Give a comprehensive overview on the existing literature to your peers
  - Clearly outline approaches to the problem and relate them to each other  
Don't just list approaches
  - Identify gaps in current research and potential future work

# Your Task for the Project

- ❑ Conduct a practical research project on a current research topic
- ❑ Write a term paper which summarizes your progress
- ❑ Deliverables
  - Code
  - Presentation (25 min presentation + 10 min Q&A)
  - Term paper (5 pages + unlimited space for references; double-column LaTeX template provided)
- ❑ What makes a good research project?
  - Research questions motivate experiments
  - Observe current baselines
  - Grounded evaluation and analysis
- ❑ What makes a good project paper/presentation?
  - Identify and motivate a problem and describe its negative impacts
  - Provide illustrative examples to clarify the problem
  - Identify gaps in current research and potential future work

# Context

- ❑ Topics are closely related to our research interests and cover a broad range
- ❑ Choose the topic that fits your own interests best
- ❑ For each topic, we suggest a type (Project/Seminar; Bachelor/Master). If a topic you prefer does not match your type, please contact us
- ❑ Diving into a research field through a seminar/project can be a good starting point for future research and theses with us
- ❑ Prerequisite: Foundational knowledge in machine learning  
e.g., as taught in our course “Machine Learning for Language Technologies” [[temir.org](http://temir.org)]

# Getting to Know Each Other

- ❑ What are you interested in?
- ❑ Have you had prior exposure to
  - machine learning,
  - natural language processing or
  - information retrieval?
- ❑ What is your scientific background?

# Topic: Mapping Abstract Concepts to Specific Instantiations

**Supervisor:** Niklas Deckers ([niklas.deckers@uni-kassel.de](mailto:niklas.deckers@uni-kassel.de))

**Suggested Type:** Project (Master)

## Task

When visualizing abstract concepts such as friendship or diligence, there exist many different instantiations (image composition, visual style, ...). This holds for image retrieval and image generation. In this project, a map between two embedding spaces will be developed that encodes the instantiation of visual representations w.r.t. abstract concepts. You will work with multimodal embeddings (such as CLIP) and generative models (such as Stable Diffusion).

# Topic: Continuous Movement within the Embedding Space of a Finite Index

**Supervisor:** Niklas Deckers (`niklas.deckers@uni-kassel.de`)

**Suggested Type:** Project (Bachelor)

## Task

Transfer lessons learned from movement within the infinite index to finite indexes: On the infinite index, arbitrarily small movements can be performed to alter image generation in any desired fashion. However, in the context of image retrieval (on a finite image dataset), such movements are inherently constrained. By defining a projection from the infinite embedding space to the finite dataset, this constraint can be lifted. Similar scenarios also exist for text when generating discrete tokens from embeddings.

# Topic: LoRA Fine-tuning for Generative Text and Vision Models

**Supervisor:** Niklas Deckers ([niklas.deckers@uni-kassel.de](mailto:niklas.deckers@uni-kassel.de))

**Suggested Type:** Seminar (Bachelor)

## Task

Compare LoRA approaches for generative text and vision models. When finetuning generative models (to generate specific types of text or a certain style of images), LoRA can be used to significantly reduce the required effort. Based on a fundamental understanding of the Transformer architecture, you will develop a comparison between the application of LoRA on image and text models.

# Topic: CycleGAN for Stable Diffusion

**Supervisor:** Niklas Deckers ([niklas.deckers@uni-kassel.de](mailto:niklas.deckers@uni-kassel.de))

**Suggested Type:** Project (Bachelor)

## Task

Find a prompt modifier that describes the relationship between a given set of image pairs (e.g., summer – winter) and ground the generation of arbitrary other images in this prompt.

# Topic: LLMs and Causality

**Supervisor:** Tim Hagen ([tim.hagen@uni-kassel.de](mailto:tim.hagen@uni-kassel.de))

**Suggested Type:** Seminar (Bachelor/Master)

## Task

Can Large Language Models (LLMs) truly understand and reason about causality, or do they simply recite patterns from their training data? Explore this topic by examining methods that use LLMs to extract causal knowledge from text, methods that extract causal knowledge from LLMs, and methods used to evaluate their causal reasoning abilities. Also investigate the existence and quality of LLM-generated causal datasets, including their usefulness compared to human-curated datasets, and summarize key findings and ongoing debates in recent research on the capabilities and limitations of LLMs in causal tasks.

# Topic: Models for Causality Extraction

**Supervisor:** Tim Hagen ([tim.hagen@uni-kassel.de](mailto:tim.hagen@uni-kassel.de))

**Suggested Type:** Seminar (Bachelor/Master), Project (Bachelor)

## Task

Causality extraction is the task of identifying cause-effect pairs in natural language text. Even before large language models (LLMs), statistical, machine learning and deep learning models have been applied to this task. The question naturally arises whether LLMs outperform more traditional approaches. Research prior work on causality extraction and identify the key ideas of the proposed algorithms and architectures and relate them to each other.

# Topic: K-CNN

**Supervisor:** Tim Hagen ([tim.hagen@uni-kassel.de](mailto:tim.hagen@uni-kassel.de))

**Suggested Type:** Project (Master)

## Task

The knowledge-oriented convolutional neural network (K-CNN) combines world-knowledge and syntactic signals to perform causality identification. Given a text with marked entities, it classifies whether one entity is claimed to have caused the other, e.g.,

The `<e1>earthquake</e1>` triggered a `<e2>flood</e2>`.

Its code and pretrained model were not published and thus cannot be compared against later approaches. Reproduce the original paper and compare K-CNN to a simple transformer baseline.

# Topic: Evaluating Agents on Structured Processes

**Supervisor:** Simon Ruth ([simon.ruth@uni-kassel.de](mailto:simon.ruth@uni-kassel.de))

**Suggested Type:** Seminar (Bachelor/Master), Project (Bachelor/Master)

## Task

LLM agents are increasingly expected to act in settings such as travel requests, booking workflows, or customer support. How can agents be evaluated in such structured, rule-governed processes? How can (1) the quality of an agents final outputs, (2) its adherence to established processes, (3) the recognition of edge cases and the involvement of humans where necessary be evaluated? Provide an overview of existing evaluation approaches and sketch one concrete process case that could be modeled, for example with BPMN (Business Process Model and Notation), as an evaluation scenario.

# Topic: FCA and Boolean Retrieval

**Supervisor:** Klara Gutekunst (`klara.gutekunst@uni-kassel.de`)

**Suggested Type:** Seminar (Bachelor)

## Task

Boolean retrieval and Formal Concept Analysis (FCA) share a set-theoretic foundation, suggesting a formal correspondence. Study how they can be mapped to each other, and analyze how FCA can model document spaces, incorporate external knowledge, structure and reduce result sets, and support ranking. Consider representations of multi-valued features and the role of FCA as an exploratory tool in information retrieval. Provide formal definitions of the mappings, illustrate them with small examples, and discuss how the use of FCA in information retrieval has evolved over time.

# Topic: Forensic Linguist Agent

**Supervisor:** Klara Gutekunst ([klara.gutekunst@uni-kassel.de](mailto:klara.gutekunst@uni-kassel.de))

**Suggested Type:** Project (Bachelor)

## Task

Authorship analysis in forensic linguistics aims to identify and analyze stylistic evidence relevant to authorship. Design, implement, and extend a modular pipeline in which a coordinating agent selects suitable linguistic analysis skills, invokes them, and aggregates their outputs into structured and inspectable results. Representative components may include lexical, punctuation-based, or syntactic analyses, as well as methods for uncertainty estimation and evidence aggregation. Evaluate the resulting system on suitable authorship datasets, compare it to simpler (non-agentic) baselines, and discuss its analytical usefulness, limitations, and extensibility in realistic forensic settings.

# Topic: Consistency and Distinctiveness of Idiolects

**Supervisor:** Klara Gutekunst (`klara.gutekunst@uni-kassel.de`)

**Suggested Type:** Project (Bachelor)

## Task

Idiolects capture individual writing style through consistent and distinctive linguistic patterns. Survey relevant work on consistency and distinctiveness in stylometry, and implement a reusable library that integrates existing methods. Apply the library to multiple datasets to analyze how author-specific features vary across genres and conditions.

# Seminar Agenda

- ❑ Give us your email address & join our Discord server (link via email)
- ❑ **Until April 24 (Week 1)** – Send us a ranked list of 3 topics that you would like to work on and your preferred day of the week for presence sessions
- ❑ **Week 3** – Tutorial session:
  - How to conduct literature research
  - How to manage a bibliography with LaTeX
  - How to do scientific writing
- ❑ **Week 6** – Structure + Abstract check
  - Confirm with your supervisor that you are on the right track
- ❑ **Week 9** – Halfway point
  - You should be halfway finished writing your term papers
  - Submit the current draft of your term paper anonymized to us
  - Everyone gets 3 other papers for review
- ❑ **Week 10** – Tutorial session on scientific presenting
- ❑ **Week 12** – Send us your slides for feedback
- ❑ **Week 15** – Presentations
- ❑ **Week 16** – Send us your final term papers and slides

# Project Agenda

- ❑ Give us your email address & join our Discord server (link via email)
- ❑ **Until April 24 (Week 1)** – Send us a ranked list of 3 topics that you would like to work on and your preferred day of the week for presence sessions
- ❑ Supervisor will get in touch
- ❑ **Week 2** – Cluster onboarding session
- ❑ **Week 3** – Tutorial session:
  - How to conduct literature research
  - How to manage a bibliography with LaTeX
  - How to do scientific writing
- ❑ Supervision with individual appointments
- ❑ **Week 10** – Tutorial session on scientific presenting
- ❑ **Week ~12** – Send us your slides for feedback
- ❑ **Week ~15** – Presentations
- ❑ **Week ~16** – Send us your final term papers, code, and slides