

# Multimodal Machine Learning Lab

Winter Semester 2026/2027

Niklas Deckers and Martin Potthast

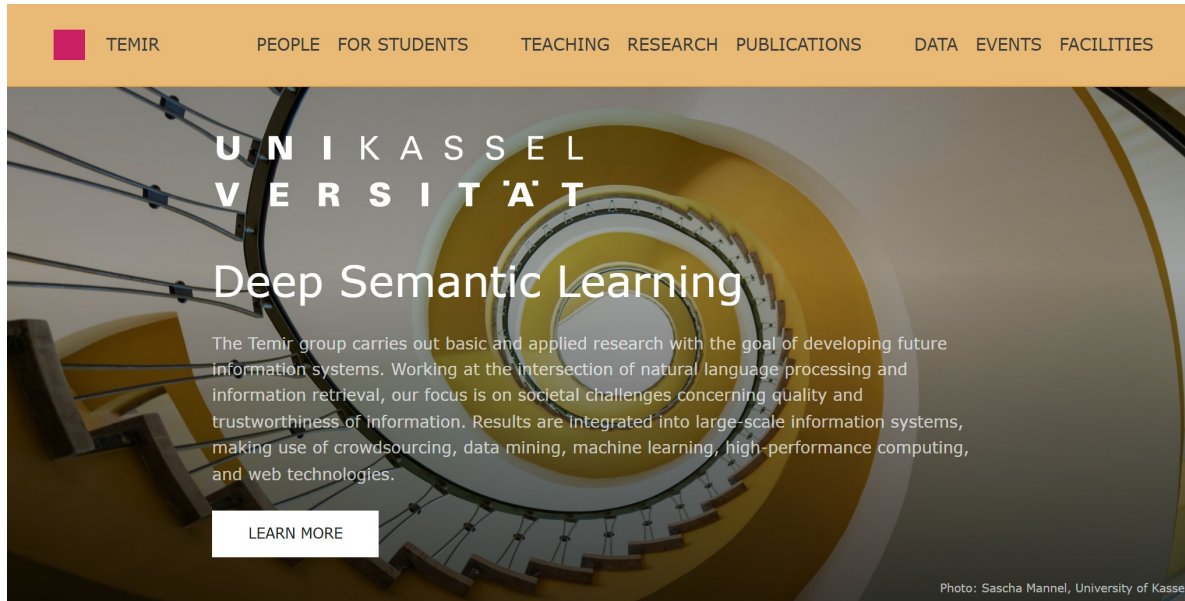
# About us



Martin Potthast



Niklas Deckers



TEMIR PEOPLE FOR STUDENTS TEACHING RESEARCH PUBLICATIONS DATA EVENTS FACILITIES

## U N I K A S S E L V E R S I T Ä T

### Deep Semantic Learning

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.

[LEARN MORE](#)

Photo: Sascha Mannel, University of Kassel

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

# Agenda

- ❑ Learning and Research Objectives
- ❑ Lab Organization
- ❑ Getting to Know Each Other
- ❑ Motivation: Multimodal Models and CLIP
- ❑ Initial Task

# Learning Objectives

- ❑ Work in a structured and self-supervised manner
- ❑ Work on a project of a larger scope
- ❑ Deal with open-ended tasks
- ❑ Groupwork and communication skills
- ❑ Apply and extend current research and tools in the field of multimodal models
- ❑ Develop and carry out experiments
- ❑ Scientific writing
- ❑ Apply machine learning to a real life problem

We would like to do real research with you!

# Research Objectives for This Lab

- ❑ How can multimodal foundation models be used in specialized applications?
- ❑ What are the capabilities and weaknesses of such models and how could these be exploited?
- ❑ How can such weaknesses be mitigated?
- ❑ Objective: Develop prototypes for multimodal applications and demonstrate attacks on the underlying models
- ❑ Using modern tools to develop ML experiments (mode depends on number of participants)

# Lab Organization

- ❑ During the semester: Project work; weekly central sessions (online? depends on number of participants) + group consultations by individual appointment
- ❑ At the end of the semester: Group presentations and written reports
- ❑ 6 CP
- ❑ Material will be provided online

# Getting to Know Each Other

What experience do you have in the following subjects?

- ☐ Machine learning, deep learning

# Getting to Know Each Other

What experience do you have in the following subjects?

- ☐ Machine learning, deep learning
- ☐ Language models like BERT

# Getting to Know Each Other

What experience do you have in the following subjects?

- ☐ Machine learning, deep learning
- ☐ Language models like BERT
- ☐ Vision models, Unets, vision transformers

# Getting to Know Each Other

What experience do you have in the following subjects?

- ☐ Machine learning, deep learning
- ☐ Language models like BERT
- ☐ Vision models, Unets, vision transformers
- ☐ CLIP

# Getting to Know Each Other

What experience do you have in the following subjects?

- ☐ Machine learning, deep learning
- ☐ Language models like BERT
- ☐ Vision models, Unets, vision transformers
- ☐ CLIP
- ☐ Programming in Python

# Getting to Know Each Other

What experience do you have in the following subjects?

- ☐ Machine learning, deep learning
- ☐ Language models like BERT
- ☐ Vision models, Unets, vision transformers
- ☐ CLIP
- ☐ Programming in Python
- ☐ PyTorch

# Getting to Know Each Other

What experience do you have in the following subjects?

- ☐ Machine learning, deep learning
- ☐ Language models like BERT
- ☐ Vision models, Unets, vision transformers
- ☐ CLIP
- ☐ Programming in Python
- ☐ PyTorch
- ☐ Git, SSH, Slurm

# Getting to Know Each Other

What experience do you have in the following subjects?

- ☐ Machine learning, deep learning
- ☐ Language models like BERT
- ☐ Vision models, Unets, vision transformers
- ☐ CLIP
- ☐ Programming in Python
- ☐ PyTorch
- ☐ Git, SSH, Slurm
- ☐ Text-to-image models like Stable Diffusion

# Getting to Know Each Other

What experience do you have in the following subjects?

- ☐ Machine learning, deep learning
- ☐ Language models like BERT
- ☐ Vision models, Unets, vision transformers
- ☐ CLIP
- ☐ Programming in Python
- ☐ PyTorch
- ☐ Git, SSH, Slurm
- ☐ Text-to-image models like Stable Diffusion
- ☐ Prompt engineering

# Foundations

- ❑ Embeddings [\[mlvu.github.io\]](https://mlvu.github.io)
- ❑ Word2Vec [\[temir.org\]](https://temir.org)
- ❑ Multimodal models aim to align different modalities (text, image, audio, ...)
- ❑ CLIP (and generative models) [\[webis.de\]](https://webis.de)
- ❑ Using CLIP as a foundation model (brainstorming: applications, vulnerabilities, mitigations)

# Initial Task

- ❑ **Until next week**
- ❑ Get on track with prerequisites: Using Python (and Pytorch), compute and compare CLIP embeddings of image–text pairs
- ❑ Search for application fields for CLIP: Find 5 exciting applications for CLIP (as precisely/actionable as possible; pairwise as different as possible)
- ❑ Later: Literature research and systematization of applications and vulnerabilities