

Multimodal Machine Learning Lab

Winter Semester 2025/2026

Niklas Deckers and Martin Potthast

About us



Martin Potthast Niklas Deckers

You can say “you” to us

TEMIR

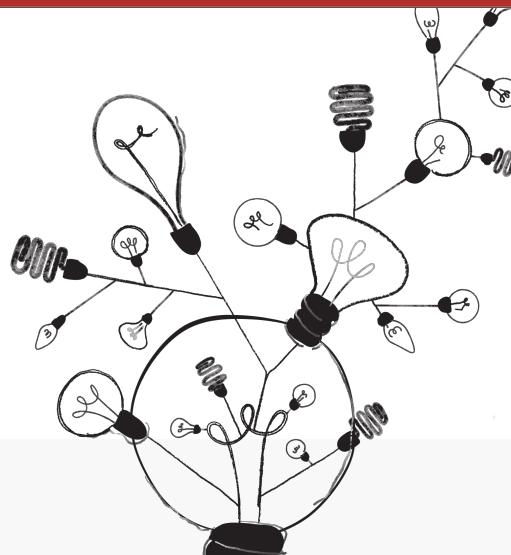
PEOPLE FOR STUDENTS TEACHING RESEARCH PUBLICATIONS DATA EVENTS FACILITIES

**UNIKASSEL
VERSITÄT**

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](#)



Search Engines Contact

Args Argument search

Deep Semantic Learning Group
University of Kassel

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

Agenda

- ❑ Motivation
- ❑ Getting to Know Each Other
- ❑ Lab Organization and Research Objectives
- ❑ Technical Foundations
- ❑ Initial Task

Introduction

Guess the movie titles:



Introduction

- Stock images: Difference between literal description and underlying meaning



Introduction

- ❑ Stock images: Difference between literal description and underlying meaning
- ❑ Also applies to the assigned tags

The screenshot shows a stock image page with the following details:

- Views:** 5.617.299
- Downloads:** 75.915
- Featured In:** Business & Work
- Location:** Erechim, RS, Brasil
- Date:** Published on October 23, 2019
- Camera:** NIKON CORPORATION, NIKON D5300
- License:** Free to use under the Unsplash License

The image thumbnail shows two young lawyers in a professional setting.

Related tags:

- Business
- Women Images & Pictures
- Work
- People Images & Pictures
- Erechim
- Rs
- Brasil
- Professional Woman
- People Images & Pictures
- Female
- Cup
- Girls Photos & Images
- Face
- Website Backgrounds
- Law
- Lawyer
- Job
- Justice
- Thermis
- Smile
- Public Domain Images

Introduction

- ❑ Stock images: Difference between literal description and underlying meaning
- ❑ Also applies to the assigned tags
- ❑ Context is relevant



Eine Abkühlung ist bald vonnöten: In Deutschland steht die erste Hitzewelle des Jahres in den Startlöchern. © IMAGO / Shotshop

Introduction

How can the following concepts be represented using generative text-to-image models?

friendship

diligence

Introduction

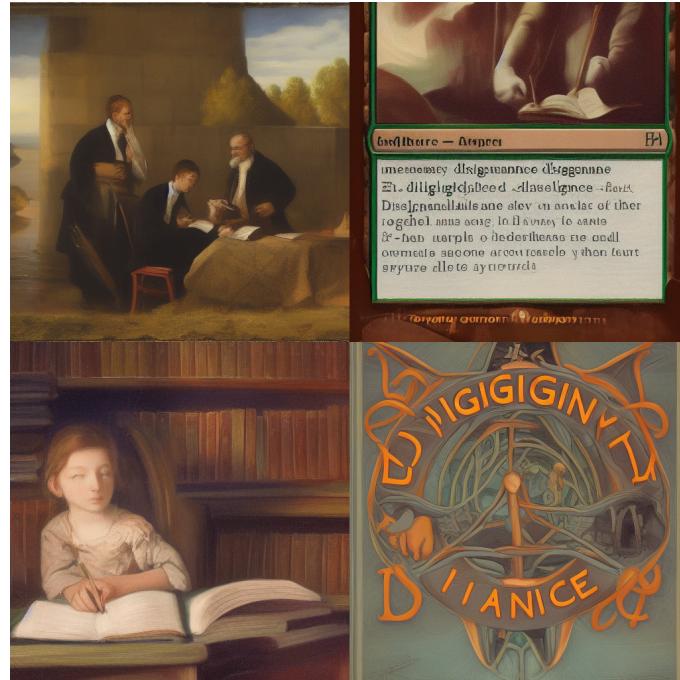
How can the following concepts be represented using generative text-to-image models?



friendship

Introduction

How can the following concepts be represented using generative text-to-image models?



diligence

Is this really the best visual representation of these concepts?

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

Lab Organization

- ❑ Weekly consultations (additional fixed time slot? where?)
- ❑ At the end of the semester: Written report (in groups) and presentation

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

- ❑ What can we learn about human–human and human–machine communication by using multimodal concepts to isolate the underlying meanings?
- ❑ What are the boundaries of this communication?
- ❑ Objective: Developing an application (web UI or game) that sets specific requirements to a communication scenario, and conducting experiments using this application

Outline of This Course

- ❑ Technical foundations
- ❑ Cluster onboarding
- ❑ Initial task: Deriving meaning from context
- ❑ Semiotics: Signs and their usage
- ❑ Communication scenarios: Real-life applications and artificially restricted scenarios
- ❑ Games as experiments
- ❑ User studies incorporating multimodality

Recap: Previous Course

- ❑ Motivation [temir.org]
- ❑ Demo [webis.de]

Foundations

- ❑ Embeddings [mlvu.github.io]
- ❑ Word2Vec [temir.org]
- ❑ Transformers [temir.org]
- ❑ CLIP and generative models [webis.de]

Initial Task



- ❑ Is there a natural order to emojis?
 - ❑ How can the semantic meaning be inferred from emojis?
 - ❑ Write a software that derives such a natural order
 - ❑ Five-minute presentations (each student individually) on 05.11.2025:
Approach, results, visualizations
 - ❑ Discussion: What ingredients are needed? What problems may arise?