Multimodal Machine Learning Lab

Winter Semester 2025/2026

Niklas Deckers and Martin Potthast

Deep Semantic Learning
University of Kassel and hessian.AI

Agenda

❑ CLIP as a Multimodal Model

❑ Homework Review: Emoji Keyboard

❑ Multimodal Scenarios

CLIP as a Multimodal Model

- CLIP and generative models [webis.de]

Homework Review: Emoji Keyboard

- 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?

Terminology

- ❑ Semiotics
- ❑ Sign, Object, Interpretant
- ❑ Icons vs. Indices vs. Symbols
- ❑ Firstness, Secondness, and Thirdness
- ❑ Signifier vs. Signified
- ❑ Sender, Intention, Message, Transmission, Noise, Receiver, Destination
- ❑ Encoding, Decoding
- ❑ Semantics vs. Syntax vs. Pragmatics

Multimodal Scenarios: Task

- ❑ Many of the following scenarios describe forms of communication. How are the steps of Encoding and Decoding represented? What are the challenges given to sender or receiver?
- ❑ How is the communication (arbitrarily) made difficult? What makes the underlying concepts difficult to describe?
- ❑ What is the role of multimodality in the given scenarios? Is it used as a hurdle or as a bridge?
- ❑ What invariances w.r.t. the context are desired? (Invariant of individual person, culture, drawing style, ...)
- ❑ Are there important or particularly interesting scenarios missing?

Multimodal Scenarios

- Image search
- Image generation
- Symbolic images in news articles
- Pictograms [imageclef.org]
- (Traffic) Signs [wikipedia.org]
- Nuclear semiotics [wikipedia.org]
- Voyager Golden Record [wikipedia.org]
- Arecibo message [wikipedia.org]

Multimodal Games

- Movie Emoji Trivia
- 4 Pics 1 Word [wikipedia.org] and generative variants [dalledle.com], [github.io]
- ESP Game [wikipedia.org]
- Quick, Draw! [quickdraw.withgoogle.com]
- Skribbl.io [skribbl.io]
- Gartic Phone [garticphone.com]
- Guess Who? [wikipedia.org]
- Scribblenauts [wikipedia.org]
- Codenames [wikipedia.org]
- Portrayal [wikipedia.org]
- Dixit [wikipedia.org]

Multimodal Datasets

❑ Web data [arxiv.org]

❑ Stock images

❑ The Noun Project [thenounproject.com]

❑ OpenMoji [openmoji.org]

❑ ARASAAC Pictograms [arasaac.org]

Mining for Abstract Concepts

- ❑ Dictionaries
- ❑ Reddit: r/captionthis [reddit.com]
- ❑ Giving game instructions for games like Unfair Mario [archive.org] or The Witness [wikipedia.org] – is this easier with words or with graphs?

Restricting the Set of Forms

- Emojis
- GIFs (in chat context)
- Toki Pona [wikipedia.org]
- Vector graphics
- Pixel art, voxel art

Next Steps

- Designing a form of Turing test
- How can we find difficult task in terms of the boundary of what humans and machines can do?
- Game design
- Bringing this into context of Semiotics literature