Multimodal Machine Learning Lab

Winter Semester 2025/2026

Niklas Deckers and Martin Potthast

Deep Semantic Learning
University of Kassel and hessian.AI

Agenda

- Literature Review
- Systematizing Multimodal Communication Scenarios
- System Design Choices

Literature Review

- Charles S. Peirce *Semiotische Schriften* and Umberto Eco *Einführung in die Semiotik* provide foundations for the field of Semiotics
- David Crow *Visible Signs* illustrates basic terms with examples from graphics design
- Paul Cobley and Litza Jansz *Semiotics for Beginners* aims for an approachable, illustrative approach in explaining the basic terms
- Thomas Friedrich and Gerhard Schweppenhäuser *Bildsemiotik* examplains basic terms of Semiotics, puts them in relation with figures of speech and gives examples from advertisements
- Sean Hall *This Means This, This Means That: A User's Guide to Semiotics* presents ambiguities that pose a challenge in the context of Semiotics; this may serve as a starting point for designing difficult tasks
- Walter Bohatsch *Typojis* presents abstract concepts that are difficult to represent in writing

Literature Review

- Felix Sockwell *Thinking in Icons* shows strategies for the design of symbols and icons
- Holger Ziemann *Icons* describes the process of designing symbols and icons with references to Semiotics
- Marcel Danesi *The Semiotics of Emoji* studies the messages sent through emojis, their design, meaning and cultural coding
- Michael Beißwenger and Steffen Pappert *Handeln mit Emojis* study the use of emojis in chat communication
- Diana Kamin *Picture-Work* describes how libraries and archives work with images (including traditional indexing and modern crawling methods) and references Semiotics
- Albert Cairo *The Functional Art* describes dimensions that are important to the design of information graphics and Sandra Rendgen and Julius Wiedemann *Information Graphics* gives examples
- Francesco Franchi *Designing News* gives case studies on how news articles and magazines are designed and illustrated to convey information

Literature Review

- Roger Fawcett-Tang *Mapping Graphic Navigational Systems* gives case studies on the design of maps including spatial and temporal information
- Beate Kling and Torsten Krüger *Spatial Orientation*, Andreas Uebele *Signage Systems and Information Graphics* and Michelle Galindo *Signage Design* give case studies on how signs can be designed to help orientation
- Eduardo Neiva *Communication Games* describes communication in a historical and societal context, but also references game theory

Systematizing Multimodal Communication Scenarios

Multimodality as a hurdle or as a bridge

- While multimodality often forms a hurdle for the interpretation (since it requires an explicit decoding step), it may be used as a bridge in a tip-of-the-tongue context (when verbally describing a concept is difficult)
- When drawing is too difficult (or tedious), providing object names instead is an example for multimodal systems helping the user (e.g., Scribblenauts)
- Experiments that use multimodality as a challenge might help to extract concepts to form bridges (by forming human intelligence tasks to build multimodal training datasets or recommender systems)
- Having multimodality as a hurdle will help to identify the limits of communication

Systematizing Multimodal Communication Scenarios

What is more difficult: Encoding or Decoding?

- The encoding step may require to simulate the decoding step
- Building challenging decoding tasks becomes more difficult
- Ambiguities (polysemes, homonyms etc.) may be used for a targeted design of difficult tasks (4 Pics 1 Word), but may also pose a challenge (Scribblenauts)
- Some steps may be encoding or decoding at the same time: For Guess Who? (similarly to Codenames), extracting shared characteristic attributes from given images can be seen as the decoding step (which could be seen to require a prior encoding step of introducing potential attributes to the images), while verbalizing these attributes can be seen as the encoding step

Systematizing Multimodal Communication Scenarios

Regarding variances

- Invariance is desired typically in the decoding step: The messages should be decoded the same way despite differences regarding the individual person, culture, drawing style
- This already needs consideration in the encoding step to eliminate ambiguities; it needs to make assumptions about the knowledge of the decoder
- E.g., the Movie Emoji Trivia game highlights invariances in cases where calquing is used to transliterate titles; test scenarios may include ambiguities such as movie titles that differ between languages

Systematizing Multimodal Communication Scenarios

Regarding variances

- However, the encoding step might also aim for variance: A variety of encodings can be produced to reflect the recipient's needs, or to challenge creativity and to identify ambiguities or shared core concepts (ESP Game)
- This depends on the task: Orientation signage aims to eliminate variance in interpretation, while games like Skribbl.io or Gartic Phone encourage "failing" by coming up with different interpretations, which is considered funny
- For the example of 4 Pics 1 Word: The encoding step requires to build variance that sufficiently narrows down the given concept, but allows to decode the term invariant of the specific representation chosen (similarly to the problem of designing stock images)
- Skills being specific to the user introduces variance (e.g., naming emotions based on images)

Systematizing Multimodal Communication Scenarios

Verification steps

- One-sided challenges (where either the encoding or decoding step can be checked using a simple algorithm) allow for an easy verification
- For more complex scenarios, an encoding-decoding roundtrip might be required for verification
- A majority vote (agreement metrics) can also be used to determine success and correct answers for non-interactive translation scenarios (such as ESP Game)
- Individual users may introduce perception biases even for simple tasks
- Defining task-specific validation criteria (that might be unknown to the user as in Portrayal)

Systematizing Multimodal Communication Scenarios

Designing difficult tasks

- Objective: Finding the limits of communication (might require compensating the learning curve)
- Competitive and cooperative scenarios can both be challenging
- Difficulty may come from speed requirements, which may also enforce simplistic encodings (e.g., *Quick, Draw!*)
- Concrete objects are easier to represent both verbally and visually than abstract objects
- Drawing itself may pose a challenge (depending on the user), which also affects how successful the decoding step is (e.g., Skribbl.io)

Systematizing Multimodal Communication Scenarios

Designing difficult tasks

- Finding out which questions to ask or which suggestions to provide in order to obtain information that is difficult to describe (active-learning style), especially in a multimodal scenario (facial composite), is an interesting interactive scenario beyond simple translation
- Balancing overspecification and overgeneralization in the given description (e.g., Portrayal or Codenames)
- Some concepts are difficult to describe (e.g., spatial relations in Portrayal)
- Orientation as an objective for signage design allows for complex scenarios

Systematizing Multimodal Communication Scenarios

Additional modalities (beyond image and text)

- Audio signals (more true to some forms of communication, but makes it more difficult to isolate pragmatic aspects)
- Temporal aspects (knowledge may change over time; users may control time)
- Spatial modality (as might be represented in a game engine)

Systematizing Multimodal Communication Scenarios

About Turing Tests

- Examples: Giving textual descriptions (Memes, Trolley Problem solutions); in-game behavior
- Advantage of reduced form sets and other strict constraints: Side aspects such as spelling, writing style can not be used as clues
- Having two replaceable players (e.g., for encoding and decoding) might be a more flexible scenario than having a human and another human plus a machine (where the latter compete against each other)
- Having someone with an outside perspective judge such an interaction as human-machine or human-human might already be sufficient (but this does not allow for targeted interventions or questions to the players)
- This might yield insight about the way humans and machines communicate

System Design Choices

- Finding the right difficulty setting (should be adjustable)
- Adjustable form set (to allow for experiments)
- Documentation and logging
- Reproducibility: Deterministic model; fixing the user intention; defining clear evaluation criteria
- Well-designed user study
- Usability (System Usability Score)
- Sensibleness of the experiments: Comparison of human and machine should be possible and fair
- Reusability: Using models and datasets for further applications

Prototyping Steps

1. Jupyter Notebook
2. Web-Based Prototype (NiceGUI etc.)
3. Separated frontend and backend
4. Game Engine

Representing Multiple Users/Agents

1. Turn-based local app (Jupyter Notebook)
2. Turn-based, but holding a session (NiceGUI)
3. Lobby-based
4. Connecting players into sessions
5. Networking using game engine
6. AI models on server or on client