Cognitive Systems
Language, Learning, & Reasoning
Welcome, Class of 2025

Sherzod Hakimov

Foundations of Computational Linguistics
Universität Potsdam

http://clp.ling.uni-potsdam.de

Foundations of CompLing / colabPotsdam

David Schlangen & Team

- conversational

- collaborative

- co-agentive

- embodied

AI

interactive

Foundations of CompLing / colabPotsdam

- conversational

- collaborative

- co-agentive

- embodied

AI

interactive

- The **clembench** evaluation framework, testing LLMs on

  - integration of information that might be distributed over various turns (backwards-looking)

  - strategic use of language / actions to get closer to goal (forward-looking)
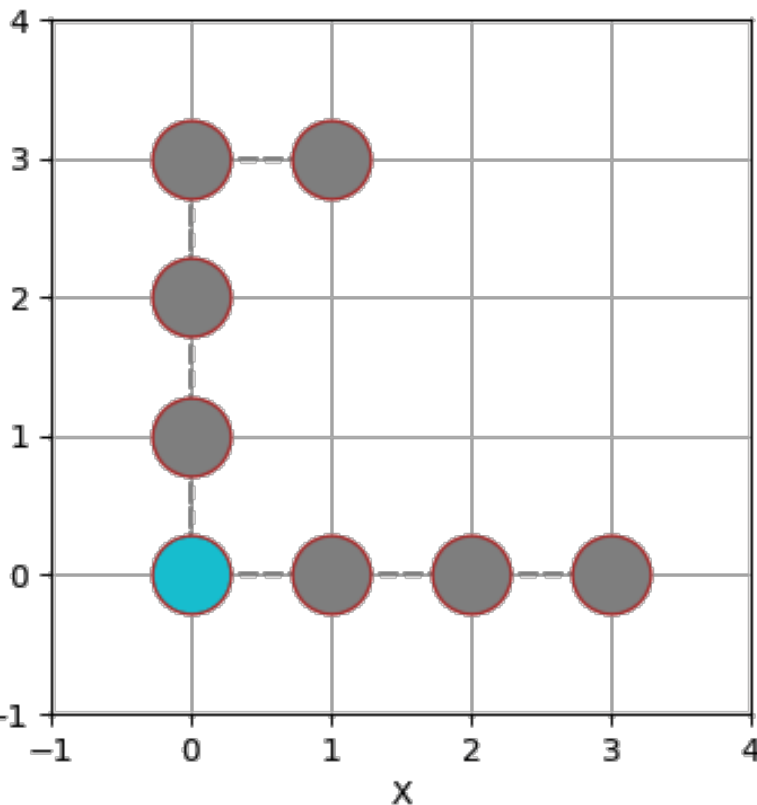
https://clembench.github.io
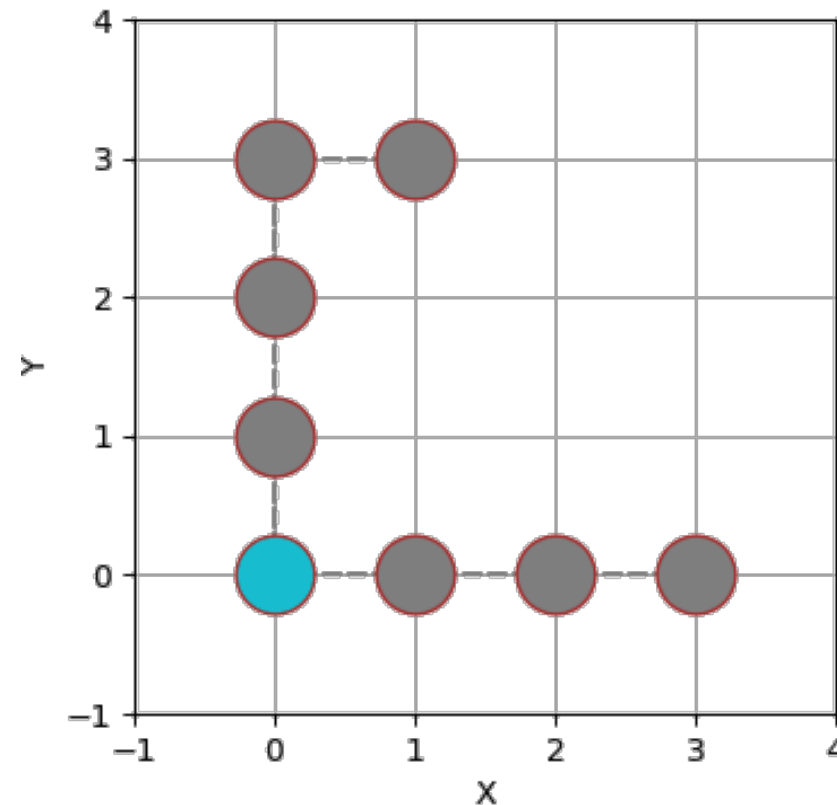
Foundations of CompLing / colabPotsdam

- conversational

- collaborative

- co-agentive

- embodied

AI interactive

- The clembench evaluation framework, testing LLMs on

  - reference game,
    image reconstruction game

  - wordle, taboo, codenames

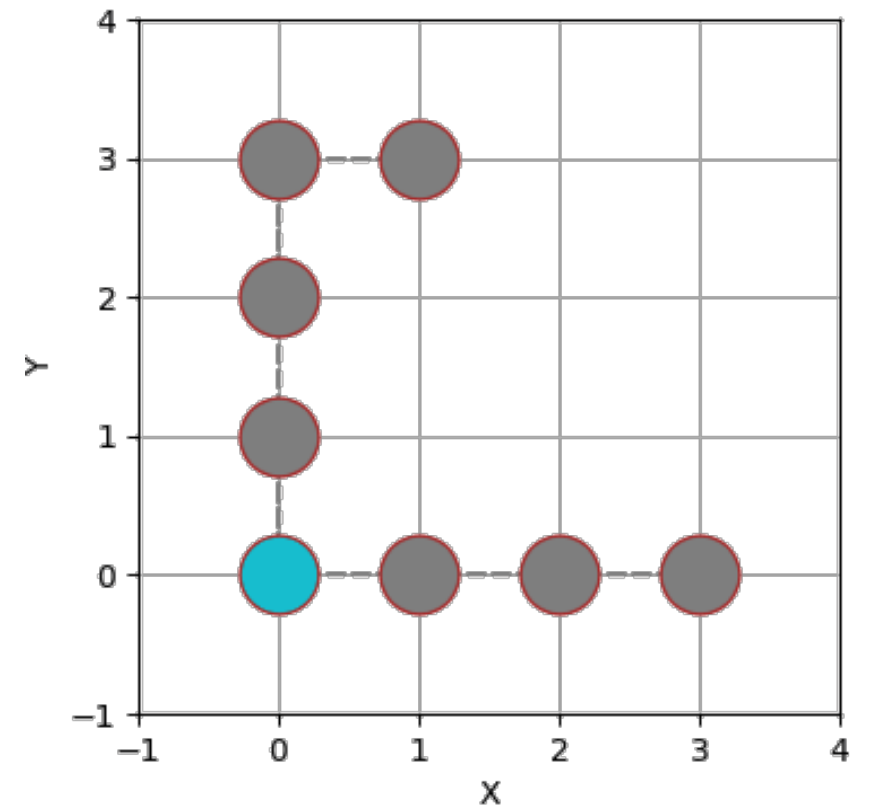 - text adventures, graph navigation

https://clembench.github.io

Foundations of CompLing / colabPotsdam

claude-3-opus   gpt-4o   Llava-1.6-vicuna-13b

Yan Weiser's 2024 CogSys MSc thesis

Foundations of Conversational

- conversatio
- collaborativ
- co-agentive
- embodied

clem-benchmark

🏆 CLEM Leaderboard

The CLEM Leaderboard aims to track, rank and evaluate current cLLMs (chat-optimized Large Language Models) with the suggested pronounciation "clems".

The benchmarking approach is described in Clembench: Using Game Play to Evaluate Chat-Optimized Language Models as Conversational Agents.

The multimodal benchmark is described in Using Game Play to Investigate Multimodal and Conversational Grounding in Large Multimodal Models

Source code for benchmarking "clems" is available here: Clembench

All generated files and results from the benchmark runs are available here: clembench-runs

CLEM Leaderboard | Multimodal CLEM Leaderboard | Plots | Versions and Details

Search for models - separate multiple queries with `;` and press ENTER...

Model	Clemscore	% Played	Quality Score
gpt-4-turbo-2024-04-09	58.3	94.88	61.45
claude-3-5-sonnet-20240620	57.08	89.64	63.68
gpt-4-0125-preview	52.5	94.92	55.31
Meta-Llama-3.1-405B-Instruct-Turbo	52.11	90.12	57.82
gpt-4-1106-preview	51.99	98.1	53
gpt-4-0613	51.09	94.88	53.85
gpt-4o-2024-05-13	48.34	85.71	56.4
gpt-4o-2024-08-06	47.71	85.71	55.66

https://clembench.github.io

Foundations of CompLing / colabPotsdam

- conversational

- collaborative

- co-agentive

- embodied

AI

interactive

- "Conversational Programming for Collaborative Robots" BMBF-funded project 2022-2024

File   Edit   View   History   Bookmarks   Tools   Help

slurk

127.0.0.1:5000   110%

Bullet Physics ExampleBrowser using OpenGL3+ [btgl] Release build
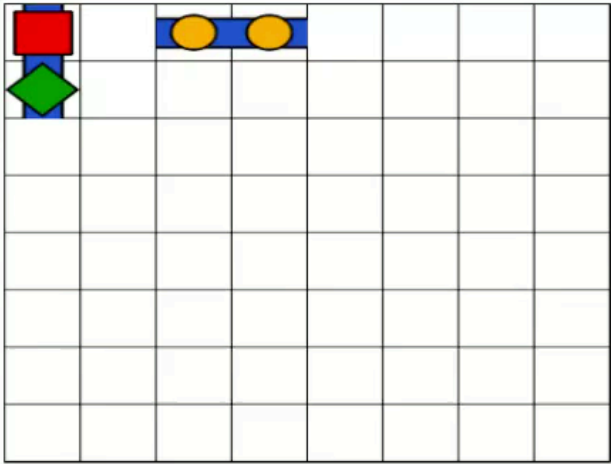
NL2Cobot Programming - DEMO

Users: ccbtsdemo, You

ccbtsdemo   14:20

Welcome to the Task Room

You can start by typing your instructions in the chat area

Read the instructions on the right ⏩

Target Board

Your task is to provide the bot with detailed step-by-step instructions to replicate the arrangement of shapes from the target board.

Available shapes

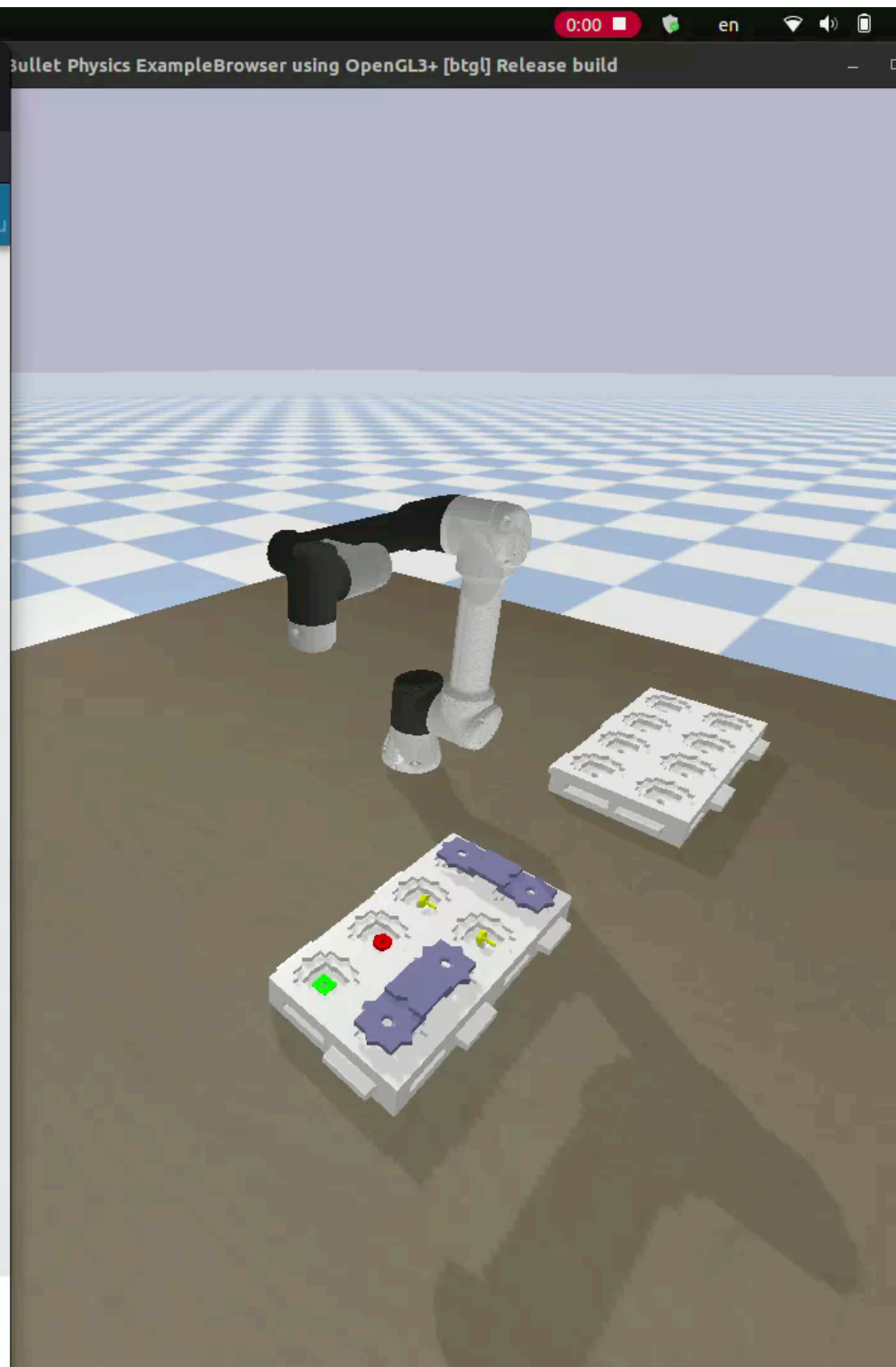◆ washer   ● screw   ■ nut   ▬ vertical bridge   ▬ horizontal bridge

Available colors:
red, green, blue, and yellow
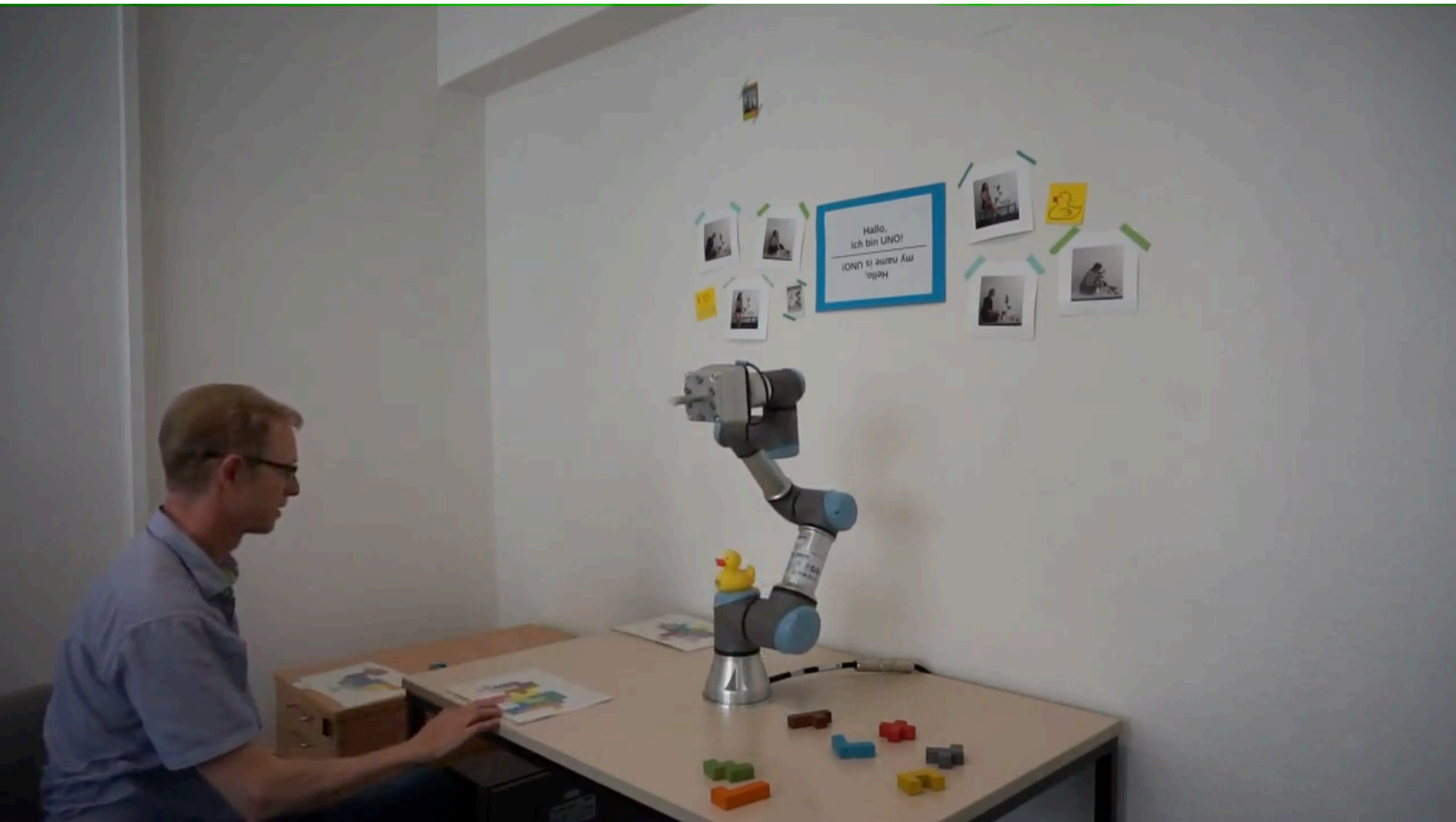
Example instructions:
1. place a red washer in the 4th row 3rd column

2. remove the last placed nut

Command to reset the target board:

Enter your message here!

Lara Pfenningschmidt's 2023 CogSys IM project

Foundations of CompLing / colabPotsdam

- conversational

- collaborative
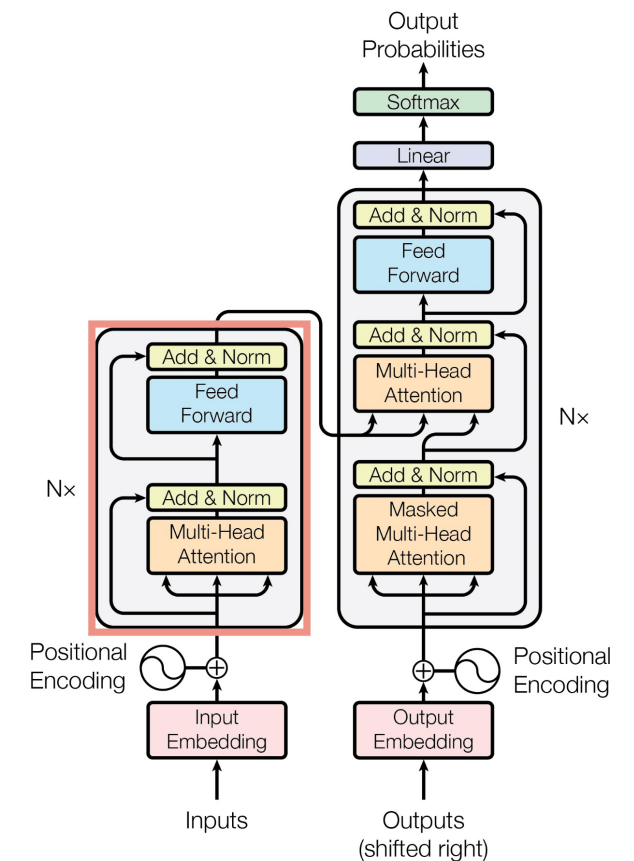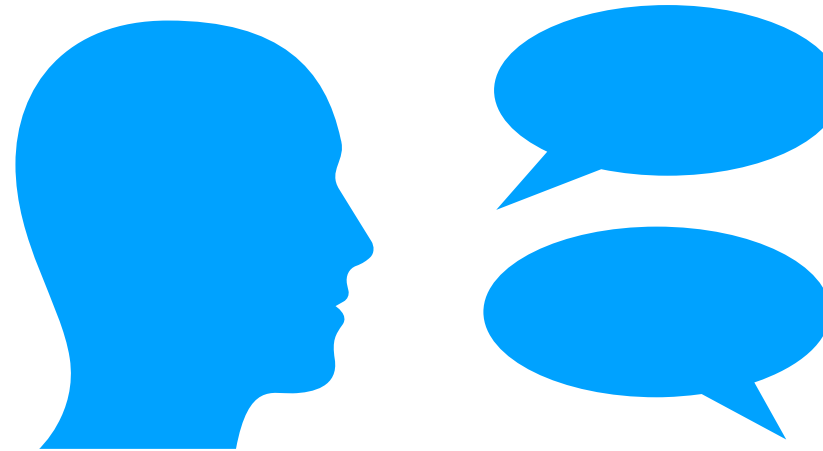
- co-agentive

- embodied

AI

interactive

Playpen

- Training LLMs on dialogue data

- Efficient way of fine-tuning using RL & SFT

HTTPS://GITHUB.COM/LM-PLAYPEN

Foundations of CompLing / colabPotsdam

- conversational

- collaborative

- co-agentive

- embodied

AI

interactive

Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Add & Norm
Masked Multi-Head Attention
N×
Positional Encoding
Output Embedding
Outputs (shifted right)

Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
N×
Positional Encoding
Input Embedding
Inputs

Summary

- Read the study regulations

- Explore widely, form your tastes

  - there is a lot of cutting edge research going on in the various labs contributing to CogSys

  - ... and in the wider context (Potsdam, Berlin)

- Enjoy — and do great research!