

## Max Planck Institute for Human Development - Center for Humans & Machines



The Center for Humans and Machines (CHM) at the Max-Planck-Institute for Human Development in Berlin conducts interdisciplinary science to understand, anticipate, and shape major disruptions from digital media and Artificial Intelligence (AI) to the way we think, learn, work, play, cooperate, and govern. Our goal is to understand how machines are shaping human society today and how they may continue to shape it in the future. The Center is composed of an interdisciplinary, international, and diverse group of scholars, and a science support team.

### **Summer Internships 2026 (voluntary)** **39,00 hours/week**

Internships will last between 6-12 weeks and start in summer 2026. This summer, we are looking for motivated student interns who are excited about working at the intersection of computer science and social sciences. The following projects are on offer:

City: Berlin; Starting date (earliest): At the earliest possible; Duration: 6 - 12 weeks;  
Remuneration: 450 €/month

#### **Tasks**

Project 1: LLM conversation ending

Our goal is to train a model that can recognize when a conversation has ended and then stop.

LLMs fall into a repetitive attractor state when used for repeated self-conversation. One theory is that this indicates a lack of self-motivated behavior on the part of the LLMs. A valid objection to this idea is that LLMs cannot indicate when they want to end a conversation. In this project, we aim to teach an LLM to produce a specialized token when it wishes to end a conversation. To do so, we need a dataset of conversation endings to train the AI to recognize an unused token that signals the end of a conversation. Then, the LLM would need to be fine-tuned without impairing its other capabilities. If successful, we will test the resulting LLM with human participants and in self-conversation to determine if attractor states persist.

- Lead Researcher: Luis Mienhardt
- Required Expertise: Machine Learning (Training or fine-tuning of an algorithm)
- Beneficial Experience: Preparing a fine-tuning dataset; fine-tuning an open-source LLM model; learning about the failure modes of LLMs; preparing small-scale LLM experiments; HPC setup
- Duration: 6 / 8 / 12 weeks

Project 2: Technical Research Assistant for AI Companion for Loneliness | Can AI companions reduce loneliness in older adults?

This project conducts a randomized controlled trial of an AI companion for older adults, offering a summer research assistant hands-on experience in interdisciplinary

psychiatry–psychology–computer science research, including data collection, data analysis, and technical support for a web interface and Python backend.

- Internship on AI Companion for Loneliness
- Lead Researcher: Rodrigo Bermudez Schettino and Chaewon Yun
- Required Expertise: Data Collection
- Additional information: Experience with web applications, Python programming and/or working with LLMs via API. Experience in data collection in behavioral experiments is a plus. Interest in online behavioral research.
- Beneficial Experience: Summer interns will get first-hand experience of interdisciplinary research between psychiatry, psychology, and computer science. Summer interns will gain experience in conducting a randomised control trial with societal relevance.
- Duration: 6 / 8 / 12 weeks

#### Project 3: (Why) do people take moral advice from Chatbots?

We are running a study on the effects of Chatbot advice on moral judgments. We have existing data from one experiment. The intern's task would be to familiarize themselves with the research question and design, analyze the data (behavioral data as well as natural language data), get familiar with new work that has been published on the topic since we ran this first study, and come up with potential follow-up experiments which we could conduct together.

- Lead Researcher: Neele Engelmann
- Required Expertise: Behavioral research methods (experimental design, data collection and analysis)
- Beneficial Experience: Experience with analyzing natural language data
- Duration: 6 / 8 / 12 weeks

#### Project 4: Participatory AI Governance: Collective Artificial Personalities

Contemporary AI systems are predominantly aligned through centralized, vendor-driven pipelines, resulting in homogenized normative profiles that reflect few of the communities they serve. This project investigates whether community-grounded governance, including shared constitutions and structured feedback, can enable AI systems to credibly represent group identities. It focuses on the design, implementation, and empirical evaluation of participatory mechanisms that allow communities to shape model behavior.

- Lead Researcher: Levin Brinkmann
- Responsibilities: Recruit and onboard partner communities, co-design constitutions and governance workflows, organize structured deliberation, curate and analyze qualitative and quantitative data, and translate community practices into formal model inputs.
- Required: Familiarity with qualitative or quantitative research methods; strong communication skills for working with community stakeholders.
- Beneficial: Background in political organizations, social movements, or cultural groups; prior experience in user research, governance design, or sociotechnical evaluation.
- You will gain: Hands-on experience in participatory governance design, sociotechnical evaluation, and translating collective social practices into formal AI alignment mechanisms, alongside ML researchers and developers.

- Duration: 8 to 12 weeks

#### Project 5: Deep Learning Intern: Collective Artificial Personalities

Post-training methods play a central role in shaping the behavioral profiles of large language models, including their response style, normative orientation, tone, and emotional expression. This project investigates how collective personalities can be encoded, stabilized, and systematically evaluated within LLMs using participatory post-training pipelines. Particular emphasis is placed on multi-persona architectures and training procedures designed to achieve the representational alignment of specific groups.

- Lead Researcher: Levin Brinkmann
- Responsibilities: Build and maintain supervised and preference-based fine-tuning pipelines, implement adapter or soft-prompt personalization methods, design benchmarks for collective alignment, and evaluate model behavior across normative, linguistic, and stylistic dimensions.
- Required: Proficiency in Python and deep learning frameworks (e.g. PyTorch, HF ecosystem); experience with large-scale model training or fine-tuning.
- Beneficial: Prior work with parameter-efficient fine-tuning (LoRA, adapters, soft prompts) or preference optimization (DPO, RLHF); interest in sociotechnical AI alignment.
- You will gain: Hands-on experience with large-scale training infrastructure (H200/B200 GPUs), multi-persona post-training architectures, and community-grounded evaluation benchmark design.

Duration: 8 to 12 weeks

#### Project 6: Deep Learning Intern: Reinforcement Learning for Artificial Pedagogy

Optimizing AI systems for immediate task helpfulness often produces answer-giving behaviors that inhibit long-term student learning. This project investigates how pedagogical strategies can be discovered and optimized by shifting the reinforcement objective from task success to student generalization. Using a teacher-student framework, it examines how teacher models evolve when rewarded for the transfer performance of a less-capable student model on unseen, structurally related tasks.

- Lead Researcher: Levin Brinkmann
- Responsibilities: Develop RL pipelines (PPO, DPO, GRPO) for teacher-student interaction loops, design reward functions capturing instructional transfer, run experiments across model scales as a proxy for learner ability, and analyze emergent pedagogical strategies.
- Required: Machine learning experience including model training or fine-tuning; proficiency in Python and RL or post-training frameworks.
- Beneficial: Prior work with RLHF, reward modeling, or preference optimization; experience with large-scale post-training system design; interest in cognitive science or educational psychology.
- You will gain: Advanced training in RL for language models, reward design for educational objectives, and the study of emergent instructional behavior in human-AI systems, with access to large-scale compute (H200/B200 GPUs).
- Duration: 8 to 12 weeks

## Requirements

- Availability at 39 hours/week for the duration of the internship (depending on the project)
- Currently enrolled in a program related to the project(s)
- Attendance at the institute in Berlin (no remote working)

### **What we offer**

- Get hands-on experience with research on human-AI interaction and scientific processes
- Close collaboration with CHM researchers
- Possibility of transitioning to position of Student Research Assistant (only possible for students at a German university)
- Salary: 450€ / month (39 h/week)

### **Application**

There is no fixed application deadline. Applications will be reviewed on a rolling basis, and positions may be filled as suitable candidates are identified. Early applications are strongly encouraged.

Please note that applicants who require a visa or work authorization to intern in Germany should apply well in advance, as the visa process may take several months.

Please apply, including a CV without a photo and your latest transcript.

The data protection declaration for the processing of personal data within the scope of your application can be found here: [https://www.mpib-berlin.mpg.de/1589569/en\\_infos\\_bewerbung.pdf](https://www.mpib-berlin.mpg.de/1589569/en_infos_bewerbung.pdf)

Diversity and severe disability:

The Max Planck Society strives for gender equality and diversity, because diversity, equality, and inclusion enrich our community and promote scientific excellence. We have therefore set ourselves the goal of increasing the proportion of women in areas where they are underrepresented and employing more people with severe disabilities. With this in mind, we expressly welcome applications from people who are often underrepresented in the workplace due to characteristics such as gender identity, disability, religion, ethnicity, and age. Our website gives you an impression of how we understand and live diversity and what opportunities we offer to respond to your individual needs: [www.mpib-berlin.mpg.de/diversity](http://www.mpib-berlin.mpg.de/diversity)

More information at <https://stellenticket.de/202484/BEUTH/>  
Offer visible until 17/04/26

