

## Bachelor's / Master's Thesis (m/w/d)

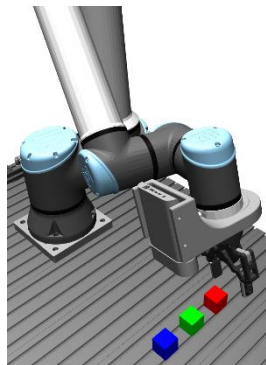
# Foundation Models in Explainable Robotics: Autonomously Extracting Robot-Internal Information for Explanations

### Problem formulation

Intuitive **Human-Robot Interaction** requires robots to reason about their internal states and decision-making processes and to **provide explanations** for their actions in a trustworthy and understandable way. Modern robots increasingly rely on **learned models** for perception and control, which often behave as black boxes, making it difficult to understand why decisions are made. At the same time, a variety of **explainable AI** techniques, as well as robot-internal information sources—such as sensor streams, logs, joint configurations, trajectories, and task histories, could potentially be used to provide insight into robot behavior. **Foundation Models**, with their language understanding and reasoning capabilities, offer a promising avenue for orchestrating such information and exploring how explanations can be generated in a flexible and adaptive manner.

### Task definition

This thesis will implement a **robotic manipulation task** where the robot relies on learned models to perform **autonomous actions**. The focus will be on investigating how internal robot states and learned components can be leveraged together with **explainable AI** methods to support explanation, and how **Foundation Models** might be applied to coordinate and synthesize this information. The effectiveness of this approach will be evaluated through experiments in which the robot executes the task and responds interactively to user questions with context-dependent explanations.



### You shall offer

- Solid knowledge base and experience in deep learning, and robotics.
- Coding skills in Python. Experience with Foundation Models, robot simulation and xAI is a plus.

### We will offer

- The chance to contribute to cutting-edge research.
- Working with state-of-the-art technology.
- Tight support from supervisors, including a workshop on scientific writing.

### Research area:

AI & Robotics

### Focus:

- ☒ Experimental
- ☐ Theoretical
- ☒ Practical
- ☒ Simulation
- ☐ Construction (CAD)

### Study program:

- ☒ Maschinenbau
- ☒ Mechatronik
- ☒ Elektrotechnik
- ☒ Informatik
- ☐ Informationswirtschaft
- ☐ Wirtschaftsingenieurwesen

**Begin:** From now on

If you are interested, please send us an e-mail with your **curriculum vitae** and a current **transcript of records**.

### Contact person:

Loris Schneider  
Geb. 50.38; Raum 1.15  
Phone: Tel.+49 721 608 48656  
loris.schneider@kit.edu

Please note that your data will be treated in accordance with the applicable data protection regulations as part of the application process.