

## Project: Simulation-Based Dataset Generation for Vision-Language-Action Models in Human-Robot Handover Tasks

Supervisors: Maj Stenmark, Esranur Ertürk

Large multimodal datasets are critical for training modern Vision-Language-Action (VLA) models in robotics such as Diffusion Policy, Action Chunk Transformers,  $\pi 0.5$ , Octo (...). However, very few tasks in existing datasets involve human-robot interaction because collecting real human-robot interaction data is expensive and difficult.



Figure 1: A robot handing over a surgical instrument to a human.

The goal of this project is to design a simulation environment capable of generating training data for human-robot handover tasks. The project will simulate a human agent and a robotic manipulator performing collaborative object transfer tasks while generating synchronized visual observations, language instructions, and robot action trajectories. The work could extend existing work that simulates a human giver and a robot receiver [1].

Students will implement:

- a simulation environment (e.g., Isaac Sim)
- scripted or motion-driven human agents
- evaluate the dataset for robot manipulation policies for handover (for your favorite VLA) in simulation and on a real robot.

The output will be a dataset suitable for training VLA models that map visual observations and language instructions to robot actions.

1 Yu-Wei Chao and Chris Paxton and Yu Xiang and Wei Yang and Balakumar Sundaralingam and Tao Chen and Adithyavairavan Murali and Maya Cakmak and Dieter Fox. *HandoverSim: A Simulation Framework and Benchmark for Human-to-Robot Object Handovers*. IEEE International Conference on Robotics and Automation (ICRA), 2022.

<https://github.com/NVlabs/handover-sim?tab=readme-ov-file>