



EAGER: Learning Language in Simulation for Real Robot Interaction

Cynthia Matuszek, Francis Ferraro, Don Engel (*cmat, ferraro, donengel@umbc.edu*)

Grounded language interactions are key to deploying robots in human environments, but gathering data from people in a variety of settings is a bottleneck. We collect simulated sensor data from the robot's side, including human interaction, in sophisticated simulations.



Human data: Talking and gesturing to a robot in VR, using a headset and the RIVR simulator (using a combination of Unity, ROS, and Gazebo).



Robot sensors: A split view of simulated robot sensor data. (Bottom: RGB; top: depth point cloud.)

Approach & System Design

- The Unity game engine lets us quickly build rich, varied worlds and scenarios
- ROS manages simulated sensing and actuation
- The Gazebo simulation environment renders robots from URDF spec files and provides simulated environment sensor readings
- Content can appear on a video wall and/or in inexpensive VR headsets



Learning about tasks: A robot interacting with a variety of food for the lunch-packing task in the Ask for Alfred/AI2Thor environment.

Goals and Broader Impacts

Targeted scenes are used to collect:

- Gesture, gaze, and speech by a human
- Sensor data from robot(s)

This will support human-robot interaction research in the community, but also **lower the barrier to entry for robotics research** by providing a rich, inexpensive testbed.