

Category 1: Numerical

Question 1 (1 point): How many pillows are on the sofa under the pictures?



Response: Print 6 in terminal.

Category 2: Object Reference

Question 2 (1 point): Find the clock on the TV cabinet



Response: Bounding box of the object

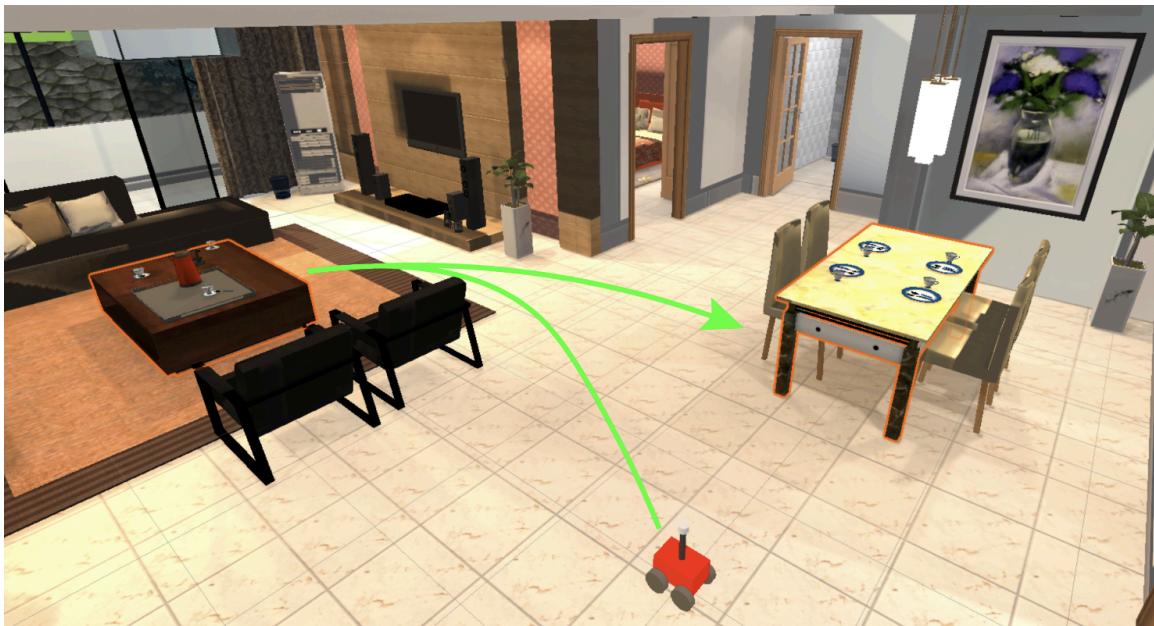
Question 3 (1 point): Find the bowl closest to the microwave on the oven.



Response: Bounding box of the object

Category 3: Instruction Following

Question 4 (2 points): Go to the coffee table with the kettle on it and stop at the dining table near the big picture.



Response: trajectory (series of waypoints)

Question 5 (3 points): First, go to the nightstand with a clock on it, then take the path between the dining table and the picture, and stop at the trash can closest to the refrigerator.



Response: trajectory (series of waypoints)

Note: The number of points of a question is based on the number of steps in the response. For example, Question 5 expects 3 steps in the response and is worth 3 points. The score is calculated as the number of correctly responded steps.