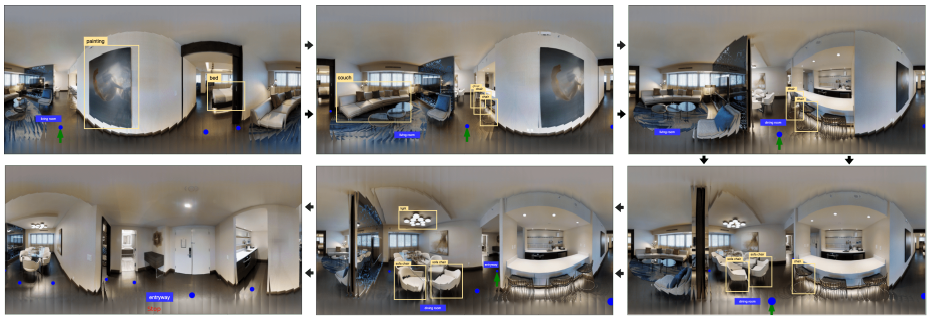
1 Supplementary material

1.1 Crafted instructions



- Walk past the large picture and chair. Walk past the dining room table turn right into the hallway and stop.

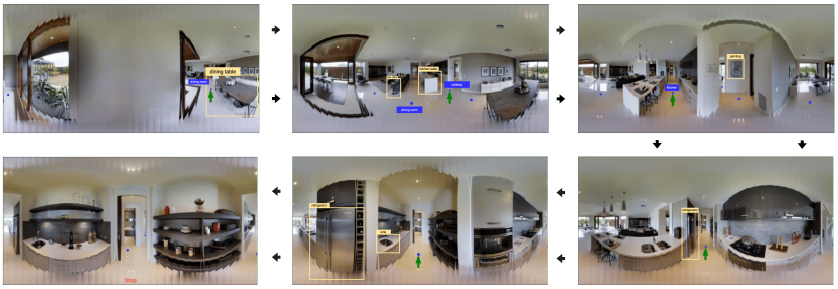
- Go straight and pass the bar with the chair/stools then pass the clear glass table with the white chairs and turn right. Wait in that hallway.

- With the painting of the mermaid towards your right, head straight. After passing the counter towards your right, turn right and wait.

- Take a left, walk straight down the left side of the painting. Walk forward. Exit the living room to the dining room. Go out of dining room into the entryway walking with the sofa chair on your left. Stop.

Fig. 1. First crafted instructions example. Panoramic views sequence above and human instructions + crafted instruction below. Images are sequenced through the arrows.

1.2 Baseline module with auxiliary tasks



- Walk passed the dining table and continue towards the kitchen. Walk through the kitchen passed the counters and stove and stop near the sink.

- Move to the far left corner of the table. Proceed to the doorway to your right.
walk into the kitchen with the stove on your right. proceed forward until the fridge is on your left
proceed forward once more until there is a small sink on your left and shelves on your right.

- Walk straight past the table and turn right to go between the kitchen counters and walk straight past the refrigerator into the pantry and stop halfway between the two shelves on the right.

- Make a right, walk straight down the left side of the dining table. exit the dining room to the hallway walking by the left side of the side table. exit the hallway to the kitchen. walk straight down the right side of the refrigerator. go straight with the sink on your left. wait at the right of the faucet

Fig. 2. Second crafted instructions example. Panoramic views sequence above and human instructions + crafted instruction below. Images are sequenced through the arrows.

1.3 Graphs on training

Fast adaption of the pre-trained agent: Figure 6 contains the pre-training with data augmentation and then finetuning starting at the best checkpoint of the pre-training. Almost five epochs are required to increase the success rate by 7% - 8%. These few epochs allow the agent to adapt to human syntax and increase the success rate.

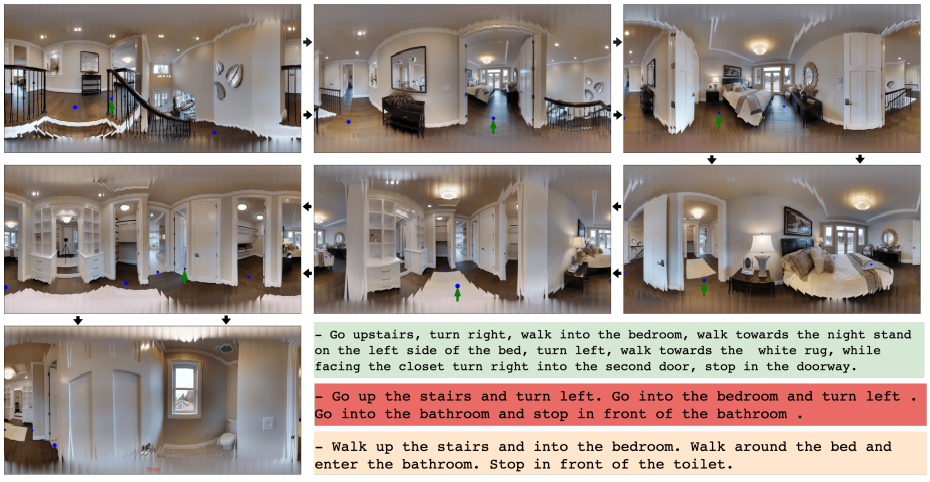
- Follow the red carpet straight and make the first left where the red carpet continues to go through the wooden doors. Wait outside of the room before entering through the wooden doors.

- Exit through double doors, walk along the red carpet to the other end of the room. Wait by the second set of double doors.

- Walk down the hallway and pass all the picture frames placed on the right until the end. Turn left to enter another room, and stop there.

- Exit the hallway to the lounge. walk straight down the right side of the chair. walk forward. go straight with the chest of drawers on your left. exit the lounge to the hallway

Fig. 3. Third crafted instructions example. Panoramic views sequence above and human instructions + crafted instruction below. Images are sequenced through the arrows.



- Go upstairs, turn right, walk into the bedroom, walk towards the night stand on the left side of the bed, turn left, walk towards the white rug, while facing the closet turn right into the second door, stop in the doorway.

- Go up the stairs and turn left. Go into the bedroom and turn left . Go into the bathroom and stop in front of the bathroom .

- Walk up the stairs and into the bedroom. Walk around the bed and enter the bathroom. Stop in front of the toilet.

- Walk up the stairs and into the bedroom. Walk past the bed and turn left. Walk into the bathroom. Stop in the doorway.

Fig. 4. First baseline module with auxiliary tasks output example. Panoramic views sequence above and human instruction + speaker follower instructions + speaker follower with objects auxiliary task + speaker follower with crafted instructions auxiliary task below. Images are sequenced through the arrows.

- Walk forward and stop at the doormat in front of the door. Turn left and continue walking. Stop at the bottom of the stairs. Walk to the top of the stairs. Then turn left and walk up the stairs. Stop walking when you're halfway up the stairs.

- Walk straight and turn left. Walk up the stairs. Stop on the second step from the top.

- Walk through the open door and then turn left. Walk up the stairs . stop on the landing at the end of the next set of stairs.

- Walk past the bathroom and turn left. Walk up the stairs. Stop on the second step from the top.

Fig. 5. Second baseline module with auxiliary tasks output example. Panoramic views sequence above and human instruction + speaker follower instructions + speaker follower with objects auxiliary task + speaker follower with crafted instructions auxiliary task below. Images are sequenced through the arrows.

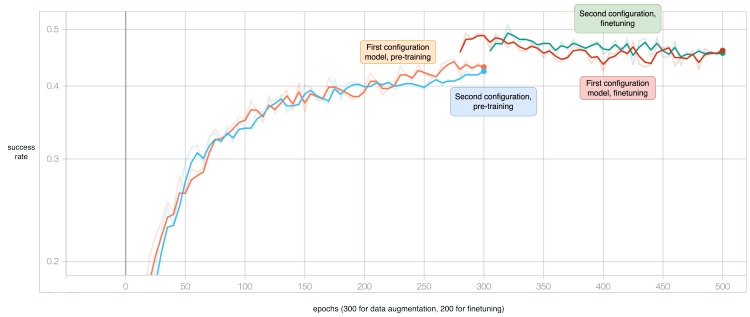


Fig. 6. Success rate on validation unseen for two different configurations. Leftmost curves (blue & orange) represent training with data augmentation (Phase I) and rightmost curves (red & green) represent fine-tuning with original training data (Phase II).