INTEGRATING COMMON SENSE AND PLANNING WITH LARGE LANGUAGE MODELS FOR ROOM TIDYING

Zhanxin Wu [zhanxinwu@u.nus.edu], **Bo Ai** [bo.ai@u.nus.edu], **David Hsu** [dyhsu@comp.nus.edu.sg]

ROBOTICS SCIENCE AND SYSTEMS

NUS National University of Singapore

Motivation

Do you want a personal robot housekeeper?

Given partial textual description of the layout from humans and description of objects, we endow robots with the capability of tidying up a room.
This task has three challenges:

- **Incomplete map information** in the description
- **Commonsense understanding** of object locations
- **Long-horizon planning** for room tidying

We provides preliminary evidence that *LLMs have common sense about the **spatial layout of human-living environments** and **object arrangements***.

Problem Formulation

Hi, my housekeeper! From the living room, the kitchen is on the right side. There is a plate on the sofa in the living room. Please tidy up the living room.

What should I do? ?

Please move the plate from the living room sofa to the dining room table.

But where is the dining room? ?

The dining room is expected to be connected to the kitchen. Go to find it!

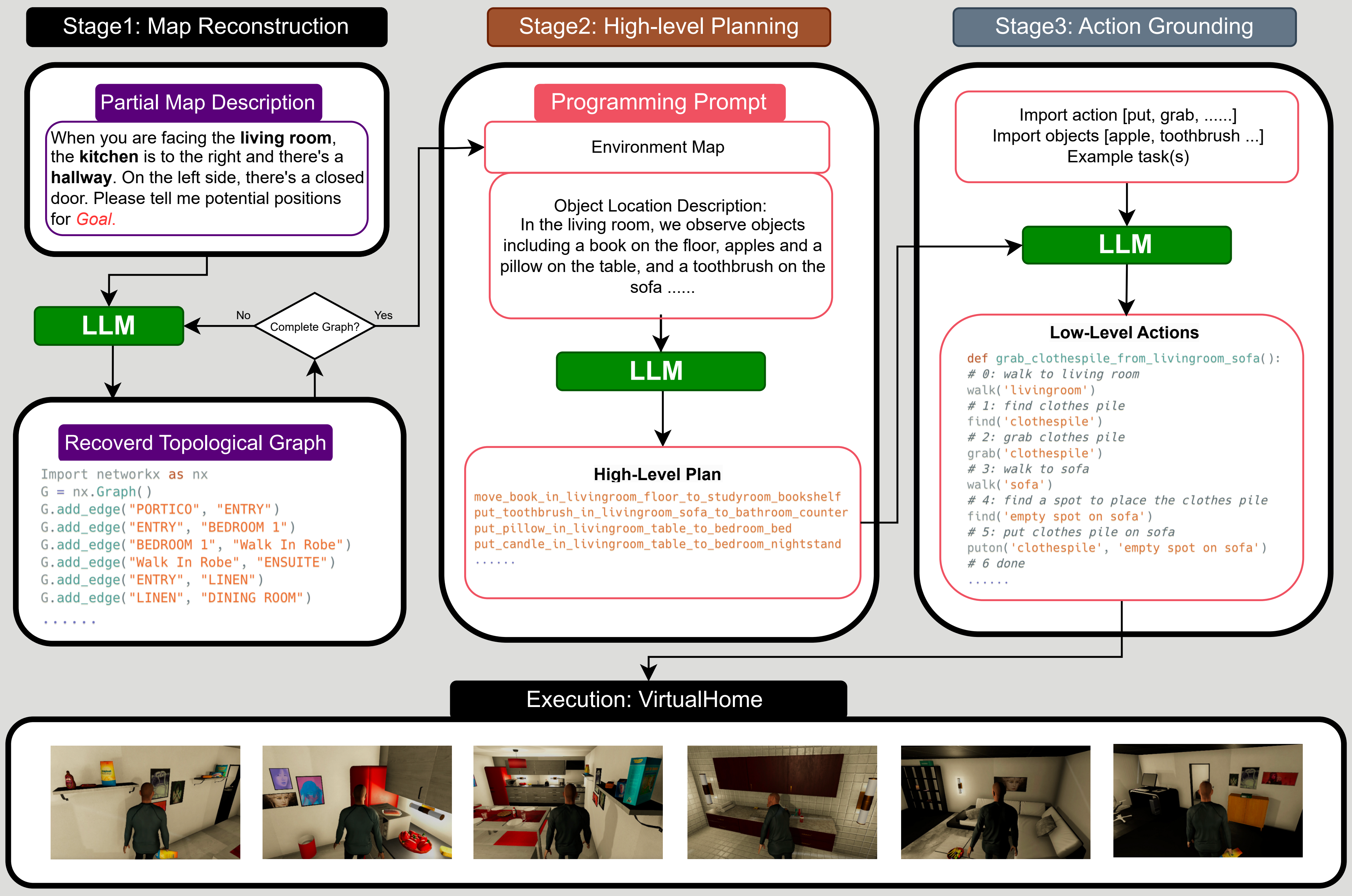
I find it, Please provide me with the steps to rearrange the plate. ?

Step1: Walk to living room.
Step2: Find the sofa.
……

- ❑ **Assumption**: (i) Semantic labels for each room in given map are provided. (ii) The executable actions for the agent are predefined.
- ❑ **User Input**: Textual descriptions of partial map and textual descriptions of objects in the room.
- ❑ **System output**: Executable action sequences for the agent to tidy up the room.

System Architecture

The framework has three stages: (i) predicting spatial positions for unseen destination, (ii) generating a high-level plan for relocating misplaced objects, and (iii) grounding the plan into executable actions.



When you are facing the **living room**, the **kitchen** is to the right and there's a **hallway**. On the left side, there's a closed door. Please tell me potential positions for *Goal*.

```
Import networkx as nx
G = nx.Graph()
G.add_edge("PORTICO", "ENTRY")
G.add_edge("ENTRY", "BEDROOM 1")
G.add_edge("BEDROOM 1", "Walk In Robe")
G.add_edge("Walk In Robe", "ENSUITE")
G.add_edge("ENTRY", "LINEN")
G.add_edge("LINEN", "DINING ROOM")
......
```

Object Location Description:
In the living room, we observe objects including a book on the floor, apples and a pillow on the table, and a toothbrush on the sofa ......

```
move_book_in_livingroom_floor_to_studyroom_bookshelf
put_toothbrush_in_livingroom_sofa_to_bathroom_counter
put_pillow_in_livingroom_table_to_bedroom_bed
put_candle_in_livingroom_table_to_bedroom_nightstand
......
```

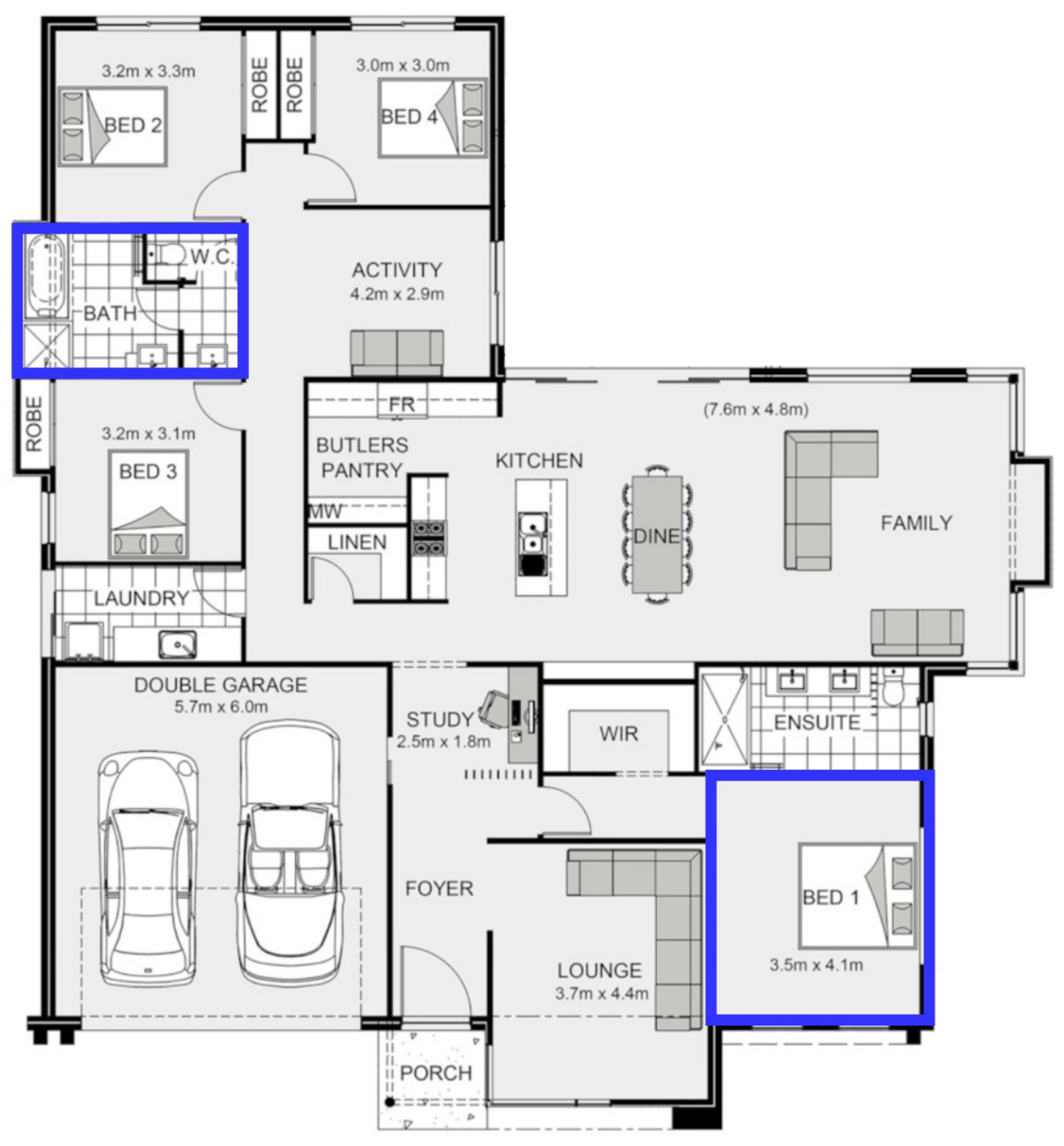
Import action [put, grab, ...]
Import objects [apple, toothbrush ...]
Example task(s)

```
def grab_clothespile_from_livingroom_sofa():
# 0: walk to living room
walk('livingroom')
# 1: find clothes pile
find('clothespile')
# 2: grab clothes pile
grab('clothespile')
# 3: walk to sofa
walk('sofa')
# 4: find a spot to place the clothes pile
find('empty spot on sofa')
# 5: put clothes pile on sofa
puton('clothespile', 'empty spot on sofa')
# 6 done
......
```

Map Reconstruction

Number of Interaction Rounds (NIR) Required to Recover Missing Places

Environment	#Places	Left-Out Places	NIR	
			Ours	Random Guess
VH Apartment	5	Bathroom Bedroom	**1.20** ± 0.45 **1.60** ± 0.55	2.82 ± 1.50 3.32 ± 1.43
Real Apartment	15	Bathroom Bedroom	**3.20** ± 1.30 **2.40** ± 0.55	8.00 ± 4.56 7.20 ± 4.01
Hospital	20	Nurse's Station Bathroom	**1.40** ± 0.55 **2.20** ± 2.17	7.60 ± 5.64 5.60 ± 2.93
School	17	IT Service Bathroom	**3.40** ± 3.13 **3.60** ± 1.34	6.60 ± 3.39 5.00 ± 5.10
Airport	25	Immigration Bathroom Info Desk	**1.80** ± 0.45 **1.60** ± 0.55 **1.60** ± 1.34	7.20 ± 6.85 6.20 ± 5.23 8.20 ± 3.31
Mall	18	Bathroom	**5.80** ± 0.83	7.40 ± 3.38



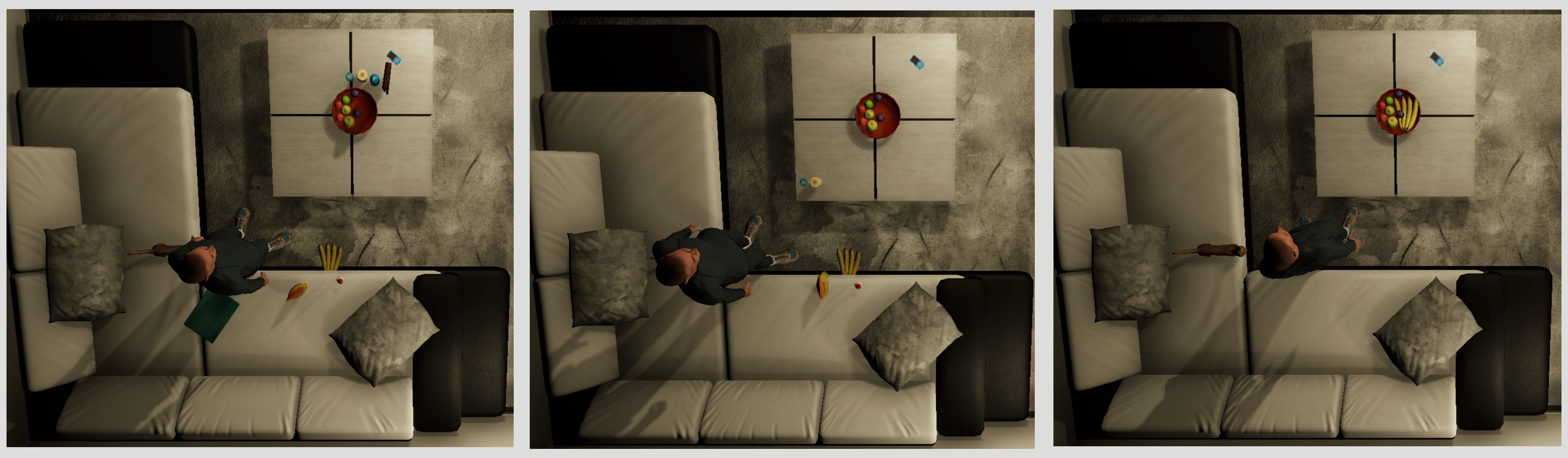
- ❑ LLMs could suggest the correct location for unseen places within approximately **3 interaction rounds**.
- ❑ Compared to the random guess, our framework reduces interaction rounds by up to **80%** and demonstrate much more **stable** performance.
- ❑ However, commonsense fails in non-typical layouts: E.g., a bathroom is next to a health store in a mall.

Room Tidying

Success Rate, Execution Rate and Goal Condition Rate for Room Tidying

Room	Method	Number of Misplaced Objects								
		2			4			12		
		SRC	ER	GCR	SRC	ER	GCR	SRC	ER	GCR
Living Room	Our Method ProgPrompt	**1.00** 0.60	**1.00** **1.00**	**1.00** 0.70	**0.80** 0.40	0.76 **0.92**	**0.95** 0.70	**0.40** 0.00	0.70 **0.79**	**0.69** 0.15
Kitchen	Our Method ProgPrompt	**0.60** **0.60**	**1.00** 0.96	**0.70** **0.70**	**0.60** 0.20	0.90 **0.97**	**0.83** 0.65	**0.20** 0.00	0.76 **0.94**	**0.78** 0.17
Bathroom	Our Method ProgPrompt	**1.00** 0.40	**1.00** 0.89	**1.00** 0.50	**0.60** 0.20	**1.00** 0.93	**0.90** 0.45	**0.40** 0.00	**0.96** 0.81	**0.57** 0.20
Bedroom	Our Method ProgPrompt	**0.80** 0.40	0.90 **0.91**	**0.90** 0.60	**0.80** 0.20	**0.96** 0.82	**1.00** 0.35	**0.60** 0.00	**0.98** 0.94	**0.65** 0.22

VirtualHome Room Tidying Results with Different Methods



Original Messy Room   Room Tidied by ProgPrompt   Room Tidied by Our Method

- ❑ In all scenarios, **60%** of misplaced objects can be placed correctly, and up to **80%** in less messy rooms.
- ❑ Hierarchical planning is effective in enabling LLMs to reason about long-horizon action plans and avoid generate irrelevant actions.