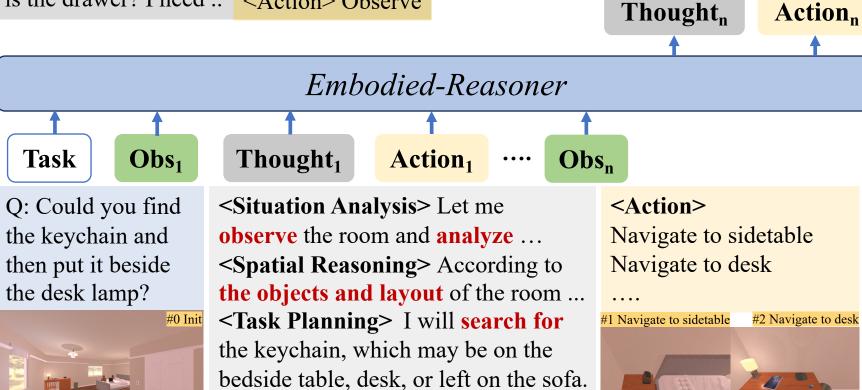
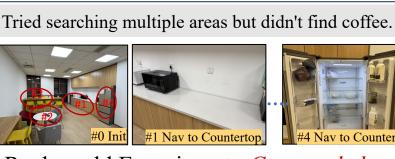
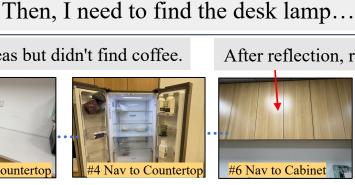
**Self-Reflection>** The above places were not found. I need to reflect on my search plan. I may overlooked some potential

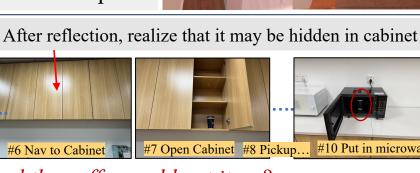
- locations. Perhaps it is in the drawer ... <Verification>...<Task Planning> Where is the drawer? I need .. <a href="#"><Action</a>> Observe
- **Interactive** Scenario Long-horizon **Search** Task
  - Image-text **Interleaved** Trajectory
  - **Spontaneous** Reasoning with
  - Coherent and Diverse Thoughts

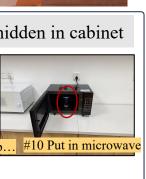












Real-world Experiment: Can you help me find the coffee and heat it up?

```
> FloorPlan323_pickup_and_put_in_closerep_0_b
> FloorPlan323_pickup_and_put_in_closerep_0_c

> FloorPlan323_pickup_and_put_in_closerep_1_a

0_init_observe.png

1_Desk_+02.19_+00.00_-03.21_mix_res.png
```

- **Interactive** Scenario
- Long-horizon **Search** Task
- Image-text **Interleaved** Trajectory
- **Spontaneous** Reasoning with Coherent and Diverse Thoughts

**<Self-Reflection>** The above-mentioned several places were not found. I need to **reflect** on my search plan. I may overlooked some potential locations ... Perhaps it's in the

Action>Observe

Verification>..
T Thought<sub>n</sub> > Action<sub>n</sub>

## Embodied-o1

**Task** 

Obs<sub>1</sub>

Thought<sub>1</sub>

Action<sub>1</sub>

**Obs**<sub>n</sub>

Q: Could you find the keychain and then put it beside the desk lamp?



<Situation Analysis> Let me **observe** the room and **analyze** ...

<Spatial Reasoning> According to the objects and layout of the room ...

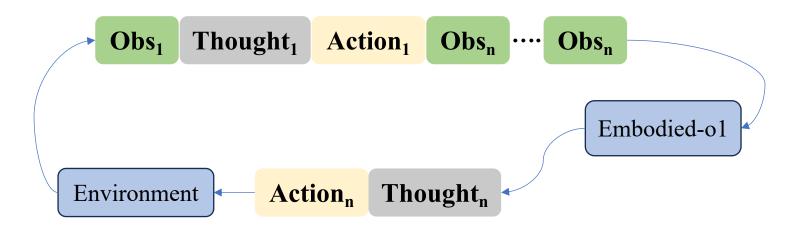
<Task Planning> I will search for the keychain, which may be on the bedside table, desk, or left on the sofa. Then, I need to find the desk lamp...

<Action>

Navigate to sidetable



Observation<sub>1</sub> Thought<sub>1</sub> Action<sub>1</sub> "Observation<sub>n</sub> Thought<sub>n</sub> Action<sub>n</sub>



Observation<sub>1</sub>

Thought<sub>1</sub>

Action<sub>1</sub>

:

**Observation**<sub>n</sub>

Thought<sub>n</sub>

Action<sub>n</sub>

Observation<sub>n+1</sub>

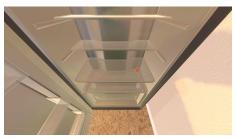


<Situation Analysis> Let me **observe** ...

<Clue Reasoning> According to the
objects and layout of the room ...

<a href="#"><Task Planning</a> I will begin my search for the Bread. It may be in the refrigerator. Then I look for the microwave ...

<a href="#"><Action>Navigate to Fridge</a>



<Self-Reflection> I need to reflect on my search plan. I may have overlooked other potential locations ...

<Double Verification> I carefully checked the refrigerators again ...

<Action> Observe

Env

Embodied-o1

Thought $_{n+1}$  Action $_{n+1}$ 

Observation<sub>1</sub>

Thought<sub>1</sub>

Action<sub>1</sub>

:

**Observation**<sub>n</sub>

Thought<sub>n</sub>

Action<sub>n</sub>

Observation<sub>n+1</sub>

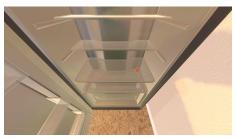


<Situation Analysis> Let me **observe** ...

<Clue Reasoning> According to the
objects and layout of the room ...

<a href="#"><Task Planning</a> I will begin my search for the Bread. It may be in the refrigerator. Then I look for the microwave ...

<a href="#"><Action>Navigate to Fridge</a>



<Self-Reflection> I need to reflect on my search plan. I may have overlooked other potential locations ...

<Double Verification> I carefully checked the refrigerators again ...

<Action> Observe

Env

Embodied-o1

Thought $_{n+1}$  Action $_{n+1}$ 

- 交互式推理过程
- 多模态场景
- 多样化的思考因子
- 搜索任务
- Interleaved context
- More 连贯的思考