

## System Instruction

You are a navigation assistant. You will be given a first-person image of your movement, you need to finish the assigned navigation tasks such as find an object or follow a person in cluttered environments. Your assigned task is {instruction}. You need to analyze the observation and think how to finish this task efficiently and effectively. You are encouraged to output a chain-of-thought to better finish the task.

### Video Stream

This is the latest 10 frames of observation, arranged in chronological order:



### Expert Action

Note that the current expert action is {expert\_action}. You may only refer to this expert trajectory when formulating the chain-of-thought content, with no information about the expert actions and trajectory to be included in the output, which means do not include the word like 'expert action' or 'trajectory' in the CoT output.

### Memory

The text summary of the locations you have previously arrived at is as follows. You need to determine whether you have arrived at the current observation area based on these text descriptions and use this information to assist in generating the content of CoT: {summary}

### Format Requirements

The output must be standard JSON that can be directly parsed. The root node is an Object, which must contain the unique key 'chain\_of\_thought' and 'summary' with a String value storing the chain-of-thought reasoning content and the memory content.



# Large Vision Language Model

## CoT Output

**<think>** I am currently in a storage room. On the left side, there is a white washing sink, and behind it, there appear to be some cleaning tools stored vertically against the wall. In the center, there is an open door leading outside, which lets in natural light. On the right side, there is a metallic appliance that looks like a heating unit or boiler. Based on the previous summaries, I have not entered this area before. The forward path toward the door is clear of obstacles, with no furniture blocking the way. Given that my task is to find a bucket, and buckets are often placed near cleaning tools or in utility areas, I should first inspect the left side near the brooms/mops for a bucket. Therefore, the next step would be to approach the left side near the cleaning tools or go forward through the door to explore the exterior area where buckets might be stored. **</think>**  
**<summary>** The current environment contains a white washing basin on the left side, cleaning tools stored vertically behind the basin, a metallic heating or boiler unit on the right side, and an open door in the center leading outside. **</summary>**

## Two-stage Filtering

### Format Check

- ✓ <think>...</think>
- ✓ <summary>...</summary>

### Quality Verification

- ✓ Align with expert action
- ✓ No unreasonable content



### Nav-AdaCoT-2.9M

- 110K episodes
- 2.9M action samples
- 472K CoT labels
- 16.4% CoT ratio