VLingNav Architecture
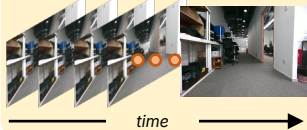
Multi-task Instruction
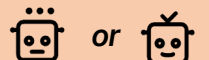- ✓ Object Goal Navigation
- ✓ Embodied Visual Tracking
- ✓ Image Goal Navigation

Observation
time

VLingNav

AdaCoT or
<think_on>
<summary>...</summary>
<think_off>

VLingMem
- 1st memory
- 2nd memory
- 3rd memory

Action Model

Robot trajectory

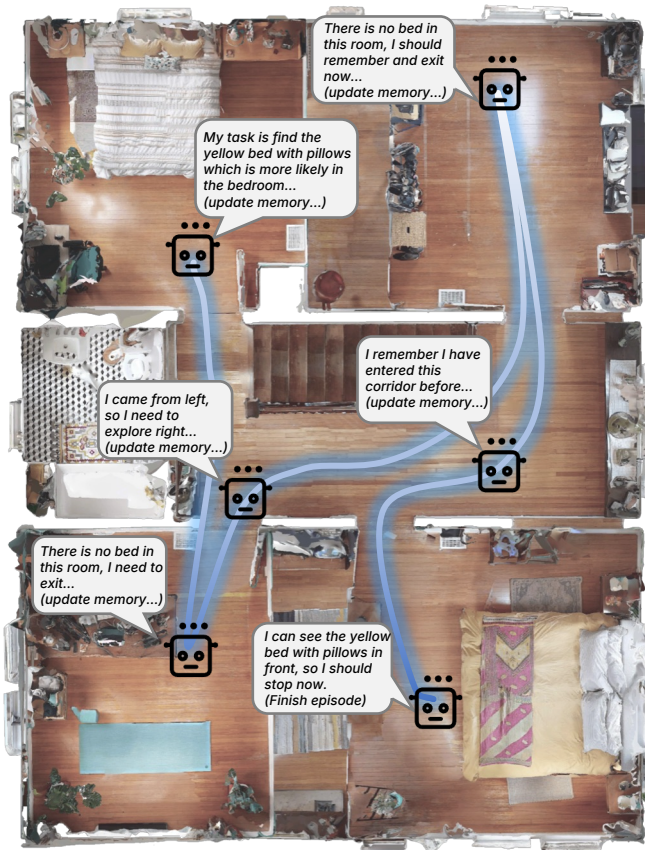## Training Recipe
Stage 1 Pre-train — VQA data
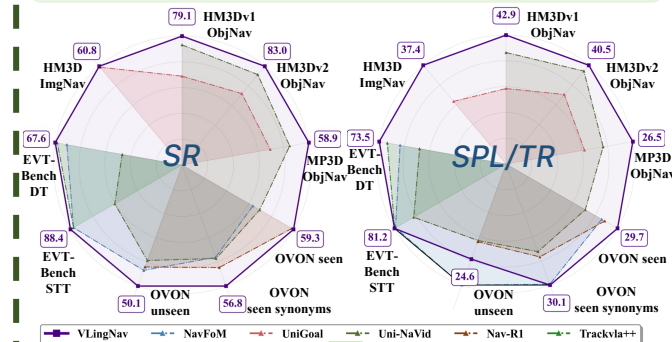**Stage 2 SFT** — Offline Nav&VQA data
Stage 3 Online RL — Online Nav data

Navigation Visualization

There is no bed in this room, I should remember and exit now... (update memory...)

My task is find the yellow bed with pillows which is more likely in the bedroom... (update memory...)

I remember I have entered this corridor before... (update memory...)

I came from left, so I need to explore right... (update memory...)

There is no bed in this room, I need to exit... (update memory...)

I can see the yellow bed with pillows in front, so I should stop now. (Finish episode)

Benchmark Performance

SR / SPL/TR radar charts

Legend: VLingNav, NavFoM, UniGoal, Uni-NaVid, Nav-R1, Trackvla++

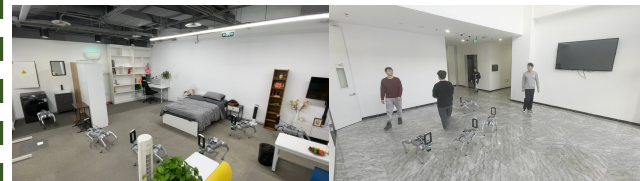# Real-world Deployment

Find the washing machine

Follow the man in red and light gray

Search for the <image_goal>

Track the man on the bike