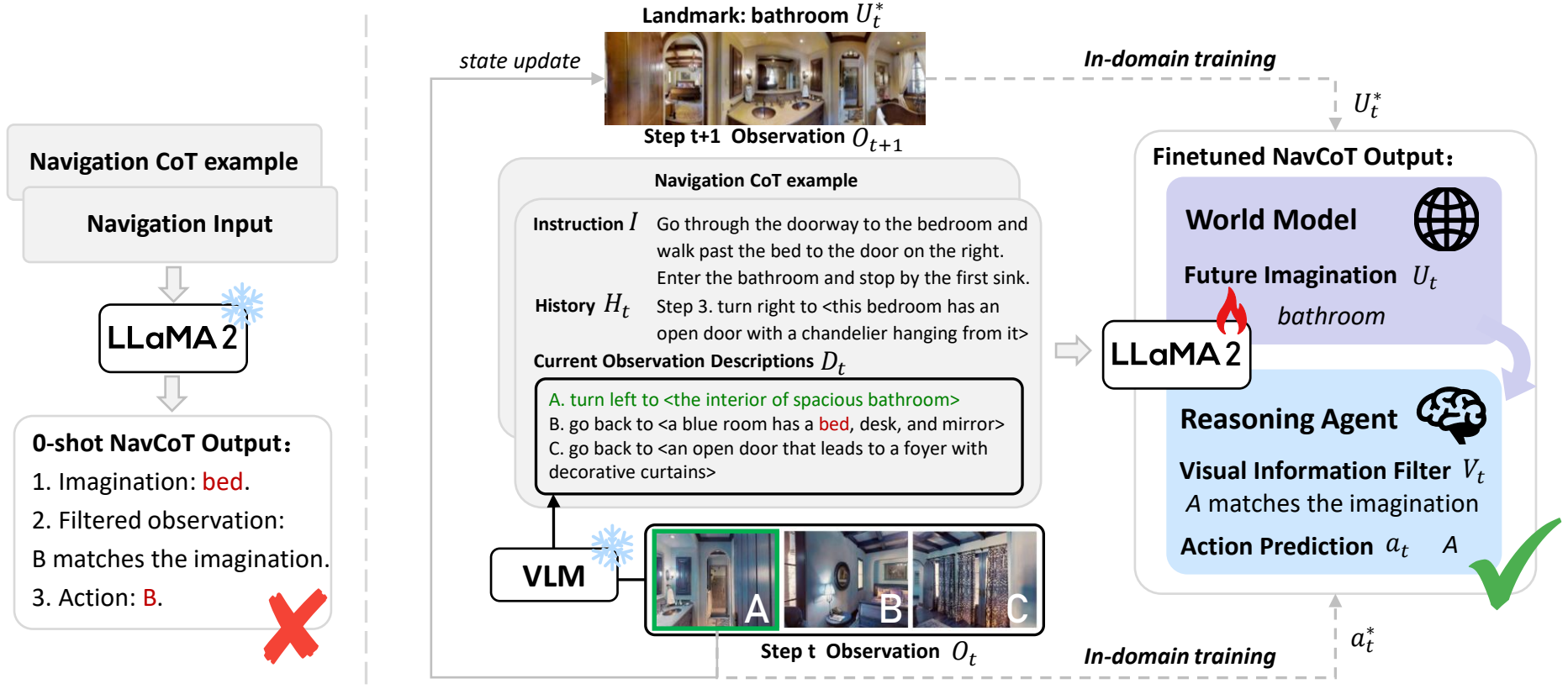
Navigation CoT example
Navigation Input
LLaMA 2
0-shot NavCoT Output:
1. Imagination: bed.
2. Filtered observation:
B matches the imagination.
3. Action: B.
Landmark: bathroom U_t^*
state update
Step t+1 Observation O_{t+1}
In-domain training
U_t^*
Navigation CoT example
Instruction I
Go through the doorway to the bedroom and walk past the bed to the door on the right. Enter the bathroom and stop by the first sink.
History H_t
Step 3. turn right to <this bedroom has an open door with a chandelier hanging from it>
Current Observation Descriptions D_t
A. turn left to <the interior of spacious bathroom>
B. go back to <a blue room has a bed, desk, and mirror>
C. go back to <an open door that leads to a foyer with decorative curtains>
VLM
A
B
C
Step t Observation O_t
In-domain training
a_t^*
LLaMA 2
Finetuned NavCoT Output:
World Model
Future Imagination U_t
bathroom
Reasoning Agent
Visual Information Filter V_t
A matches the imagination
Action Prediction a_t A