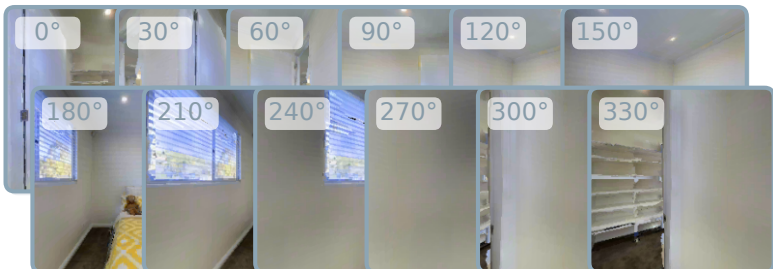


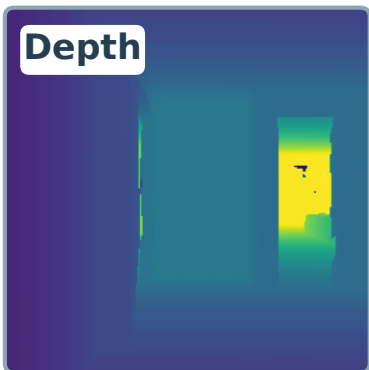
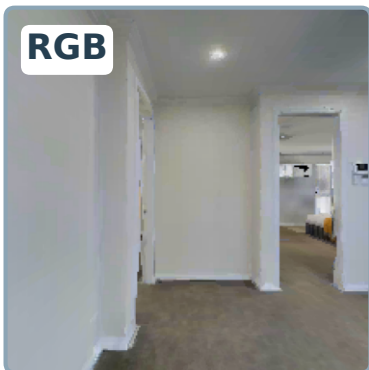
(a) Observation and context

 **Language instruction**
“Leave the bedroom and take a left in the hallway. ...”

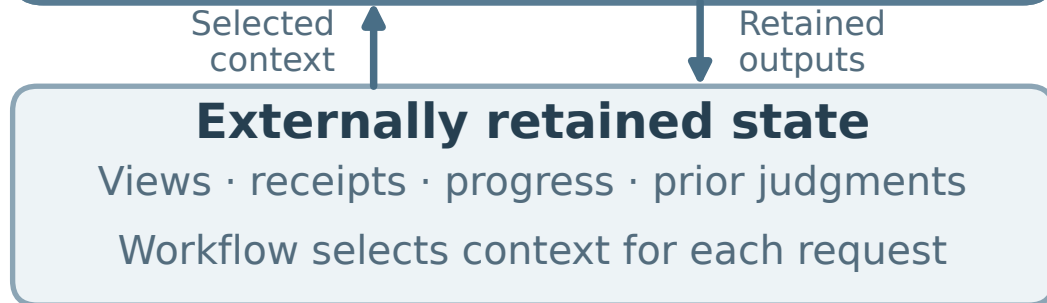
Initial panorama start only



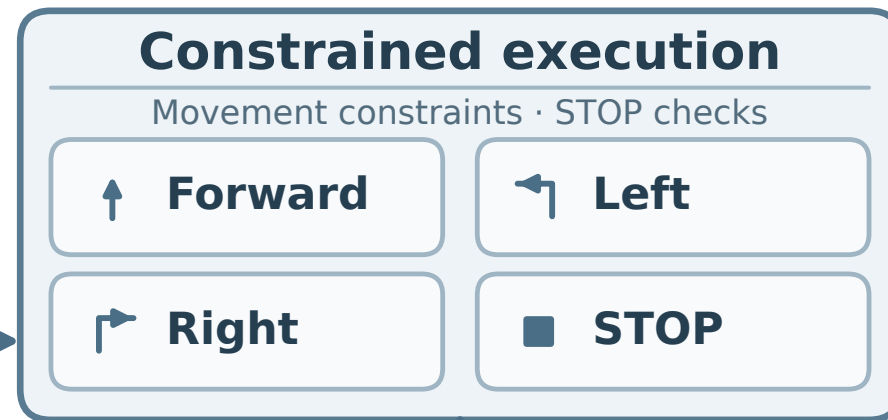
Current observation



(b) Model calls and external state

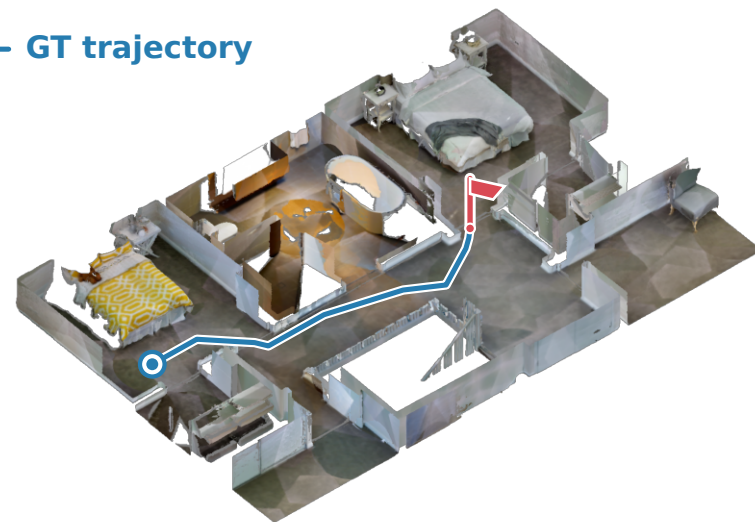


(c) Embodied interaction



execute

— GT trajectory



Environment

New observation + execution feedback