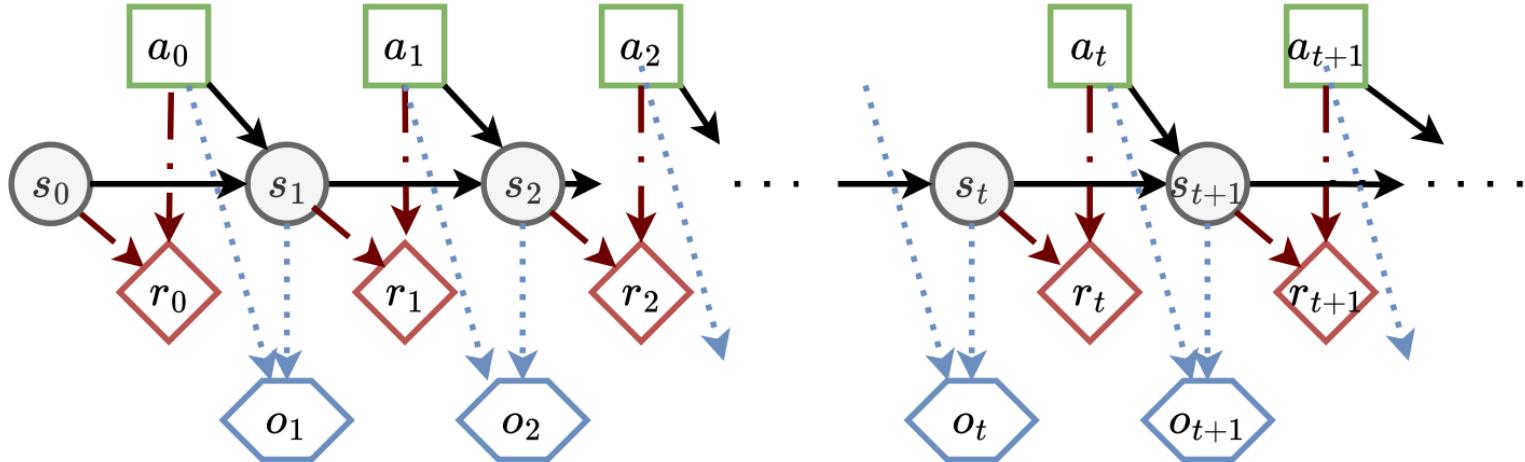


 **State:** $s_t \in \mathbb{S}$
 **Action:** $a_t \in \mathbb{A}$
 **Reward:** $r_t \in \mathbb{R}$
 **Observation:** $o_t \in \mathbb{O}$



→ **Transition Model:** $\mathbf{T} = P(s_{t+1}|s_t, a_t)$

→ **Reward Model:** $\mathbf{R} = P(r_t|s_t, a_t)$
or $\mathbf{R} = r_t(s_t, a_t)$ for deterministic reward model

→ **Observation Model:** $\mathbf{O} = P(o_t|s_t, a_{t-1})$ for Active Observation
 $\mathbf{O} = P(o_t|s_t)$ for Passive Observation