

Coach Policy Update

$$\pi_{c,t} \rightarrow \pi_{c,t+1}$$

⑨



③

**Instant
Instruction**
 I_t

Student Policy
Update

$$\pi_{l,t} \rightarrow \pi_{l,t+1}$$

⑧

⑥



②
Intended
Action
 \bar{A}_t

①
State
 S_t



⑤
Reward
 R_t

④
Action
 A_t

