



**FCAI**

**fcai.fi**

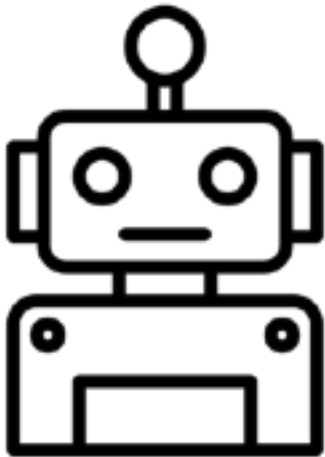


$$\max_{\pi} \mathbb{E}_{\pi, P} \left[ \sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \mid s_0 = s, \pi \right]$$

$$\max_{\pi} \mathbb{E}_{\pi, P_{\phi}} \left[ \sum_{t=0}^{\infty} \gamma^t r_{\xi}(s_t, a_t) \mid s_0 = s, \pi \right]$$

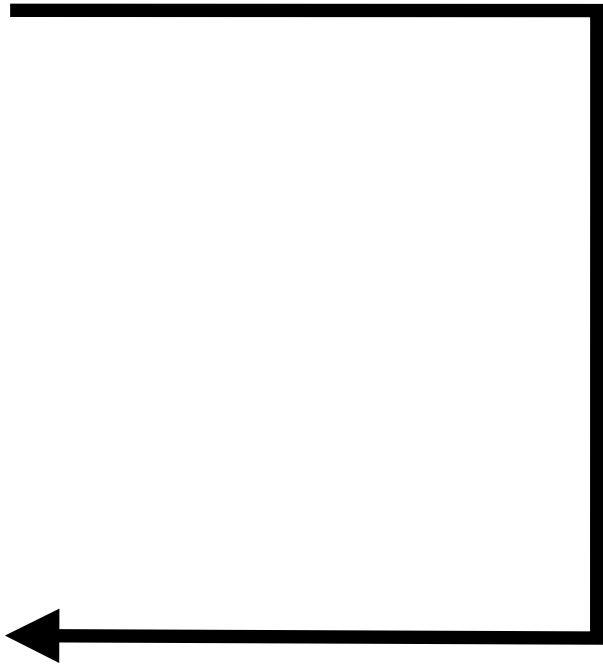
# Reinforcement Learning (RL)

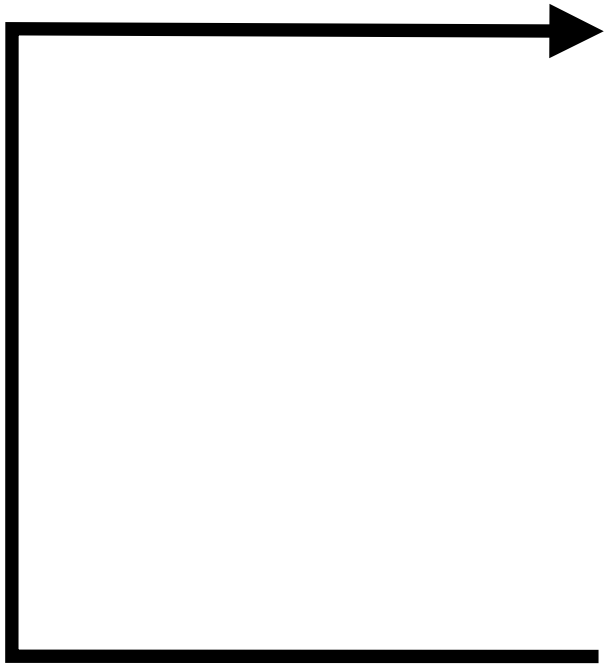












$$a_t = \pi(s_t)$$

Actions

$$s_{t+1} \sim P(\cdot | s_t, a_t)$$

Transition function

$s_{t+1}, \quad r(s_t, a_t)$

State,    Reward

States  $s \in \mathcal{S}$

Actions  $a \in \mathcal{A}$



Policy  $\pi :: \mathcal{S} \rightarrow \mathcal{A}$

Transition function  $P(s_{t+1} | s_t, a_t)$

Reward function  $r_t \equiv r(s_t, a_t)$

Discount factor  $\gamma \in [0, 1]$

# Markov Decision Process (MDP)

In model-based RL  
these are the “model”

**Goal:**

$$\max_{\pi} \mathbb{E}_{\pi, P} \left[ \sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \mid s_0 = s, \pi \right]$$

# Reinforcement Learning (RL)

## Markov Decision Process (MDP)

States  $s \in \mathcal{S}$

Actions  $a \in \mathcal{A}$

Policy  $\pi : \mathcal{S} \rightarrow \mathcal{A}$

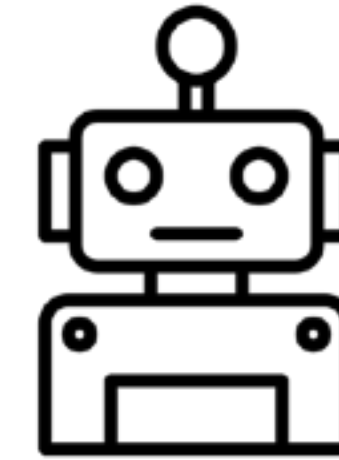
Transition function  $P(s_{t+1} \mid s_t, a_t)$

Reward function  $r_t = r(s_t, a_t)$

Discount factor  $\gamma \in [0,1]$

$s_{t+1}, \quad r(s_t, a_t)$

State, Reward



$a_t = \pi(s_t)$   
Actions

In model-based RL  
these are the “model”

$s_{t+1} \sim P(\cdot \mid s_t, a_t)$

Transition function

**Goal:**

$$\max_{\pi} \mathbb{E}_{\pi, P_{\phi}} \left[ \sum_{t=0}^{\infty} \gamma^t r_{\xi}(s_t, a_t) \mid s_0 = s, \pi \right]$$



# World Models