

# Q-Learning

Aayush Kejriwal

Adelina Akhsanova

Amin Abaza

Andres Behnsen

Dwight Alexander

Jonas Hellevang

# What is model-free RL?

Estimates value functions based on experience  
for an unknown Markov decision process

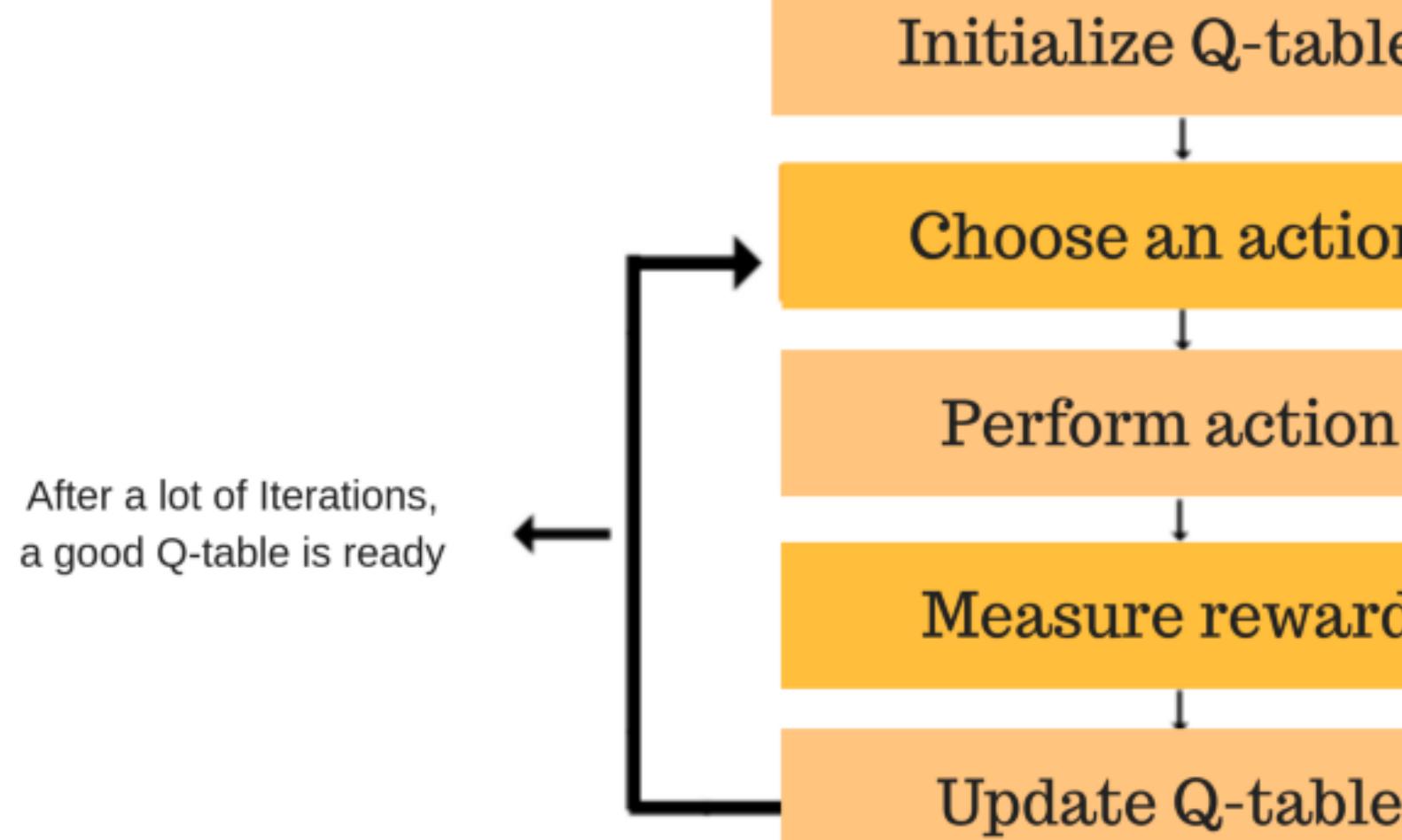


# What Is Q-Learning

**Off-policy learning**

An agent learns about how good actions are under perfect behavior  
without having to behave perfectly.

# Q-Learning Algorithm



Action	↑	↓	→	←
Start	0	0	0	0
Idle	0	0	0	0
Correct Path	0	0	0	0
Wrong Path	0	0	0	0
End	0	0	0	0

Action	↑	↓	→	←
Start	0	0	1	0
Idle	0	0	0	0
Correct Path	0	0	0	0
Wrong Path	0	0	0	0
End	0	0	0	0

Action	↑	↓	→	←
Start	0	0	1	0
Idle	0	0	0	0
Correct Path	0	50	22	0
Wrong Path	15	0	18	0
End	0	0	1	0

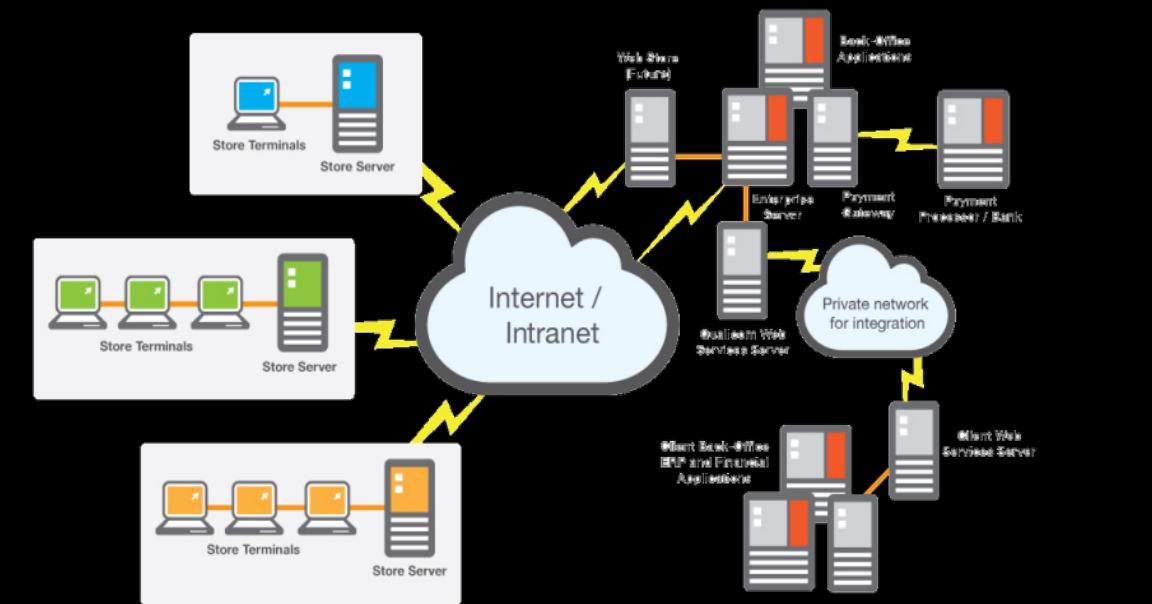
# Q-Learning Formula

Q-function uses Bellman equation and takes as inputs actions(a) and state(s):

$$Q(s, a) = Q(s, a) + \alpha [R(s, a) + \gamma * \max Q'(s', a') - Q(s, a)]$$

- New Q Value for that action and state
  - Current Q Values
  - Reward for taking that action at that state
  - Maximum expected future reward given the new state(s') and all possible actions at that new state
- 
- Learning Rate  $\alpha$  - to what extent newly acquired information overrides old information. 0 - agent learn nothing, 1 agent considers only the most recent information
  - Discount rate  $\gamma$  - significance of future rewards (0 - agent "myopic", only considers current rewards, 1 it strives for a long-term high reward.)

# Use Cases



## Web System Configuration

Autonomic reconfiguration of parameters in multi-tier web systems in dynamic environments.

## Personalized Recommendations

Using Q-values as their metric to recommend news and adding a click through rate reward to the agent.



## E-Sports/Gaming

Achieving superhuman performance in games by playing games on a loop. Learning every time.