8- deep reinforcement learning

reinforcement learning= 

rewards= positive or negative

↳ the agent's goal is to maximize the cumulative result it receives over time
↳ policy = the rules that specify how to choose an action (can be simple func, or complex nn)
↳ value function= estimates how good it is for the agent to be in a given state

types of reinforcement learning=
↳ model-based= the agent builds a model of the environment and uses it to make decisions
↳ model-free= the agent learns policies or value functions directly from experience, no model
↳ deep reinforcement learning= using deep nn to approximate policies or value funcs.

markov decision process= the probability of the next state depends only on the current state, not on previous states or actions

approaches for learning optimal policies
↳ policy gradient method= adjusting the parameters of the policy function to increase the expected cumulative result, done using gradient ascent → REINFORCE algorithm (stochastic)
↳ value-based method= choosing the action that maximizes the value function in each state (deterministic)
  ↳ Q-learning ⇒ Q-function = expected total reward from taking a specific action in a specific state
    ↳ bellman equation
      ↳ dynamic programming = create a table (# states by # actions)
  ↳ deep Q-learning ⇒ deep nn is used to approximate the Q-function

☆ reinforcement learning is an effective solution to real-world autonomous control problems