**1. Agent:**

* The learner or decision-maker in the environment. The agent takes actions based on the current state of the environment and learns from the rewards or punishments received to maximize the total cumulative reward over time.

**2. Environment:**

* The external system or world in which the agent operates. The environment provides the agent with states and rewards, and reacts to the agent's actions by transitioning to new states.

**3. State:**

* A representation of the current situation of the environment. It contains all the necessary information required by the agent to decide on an action. For example, in a chess game, the state could be the positions of all pieces on the board.

**4. Action:**

* A decision or move made by the agent. Depending on the current state, the agent selects an action from a set of possible actions. For instance, in a robotic arm, the action could be moving the arm up or down.

**5. Reward:**

* A scalar feedback signal received after taking an action in a particular state. The reward indicates the immediate benefit or cost of the action taken. The goal of the agent is to maximize the cumulative reward over time.

**6. Policy (π):**

* A policy is a strategy used by the agent to decide which action to take in each state. It can be deterministic (π(s) = a) or stochastic (π(a|s) = probability of taking action a in state s).

**7. Value Function (V):**

* The value function estimates the expected cumulative reward of a state, considering future rewards. It helps the agent determine the overall desirability of a state.

**8. Action-Value Function (Q):**

* The action-value function, or Q-function, estimates the expected cumulative reward of taking a particular action in a particular state and following a specific policy thereafter. Q-values help the agent in selecting the optimal actions.

**9. Discount Factor (γ):**

* A parameter between 0 and 1 that determines the importance of future rewards. A discount factor close to 1 values future rewards almost as much as immediate rewards, while a factor close to 0 prioritizes immediate rewards.

**10. Exploration vs. Exploitation:**

* **Exploration** is when the agent tries out new actions to discover their effects and improve its understanding of the environment. **Exploitation** is when the agent uses its current knowledge to take the best-known action to maximize rewards. Balancing exploration and exploitation is crucial for effective learning.

**11. Episode:**

* A sequence of states, actions, and rewards, ending when a terminal state (goal or failure) is reached. An episode represents a complete run from the initial state to the terminal state.

**12. Trajectory:**

* A path or sequence of states and actions taken by the agent from the beginning to the end of an episode.

**13. Markov Decision Process (MDP):**

* A mathematical framework for modeling decision-making, representing the environment in terms of states, actions, rewards, and transition probabilities. An MDP assumes that the future state depends only on the current state and action, not on past states (Markov property).

**14. Temporal Difference (TD) Learning:**

* A class of model-free learning methods that update value estimates based on the difference (temporal difference) between successive estimates. TD learning combines aspects of Monte Carlo methods and dynamic programming.

**15. Learning Rate (α):**

* A parameter that determines the extent to which new information overrides old information. It controls the speed of learning in algorithms like Q-learning.