

## Tasks

- Episodic: has starting and ending point
- Continuous: has starting but no ending point

## Exploration vs Exploitation Trade-Off

- Exploration: exploring an environment by trying random actions and receiving feedback/rewards from the environment.
- Exploitation: It's about exploiting what we know about an environment to get maximum reward.
- Exploration - Exploitation trade-off: balances how much we want to explore and exploit the environment

## Policy

- Policy - An agent's brain, tells us what action to take given state
- Optimal Policy - Policy that maximizes the reward, learned through training.

## Policy-based methods

- Approach to solving RL Problems
- Policy is learned directly, maps each state to the best corresponding action, or a probability distribution over the set of actions at that state

## Value-based methods

- Another approach to solving RL problems
- train a Value function that maps each state to the expected value of being in that state

