

## Reinforcement Learning Quiz

**Instructions:** Select the correct option(s) for each question. Some questions may have multiple correct answers. Choose all that apply when indicated.

Question

- 1) Which algorithms belong to the policy-gradient family?
- A) Proximal Policy Optimization (PPO)
  - B) Q-learning
  - C) REINFORCE
  - D) Deep Q-Network (DQN)

Question

- 2) An environment is considered partially observable when:
- A) The agent can observe the full state
  - B) Observations lack some hidden variables
  - C) The reward signal is stochastic
  - D) Transition dynamics are deterministic

Q

- 3) In Q-learning, the key difference from SARSA is:
- A) On-policy vs. off-policy nature
  - B) Use of neural networks
  - C) Continuous action space support
  - D) Model-based planning

Questi

- 4) The Advantage term in Actor-Critic methods is calculated as:
- A) State-value minus action-value
  - B) Action-value minus state-value
  - C) Reward minus entropy
  - D) Policy-gradient estimate

Question

- 5) Techniques to stabilize deep-RL training include:
- A) Experience replay
  - B) Target networks

- C) Layer normalization
- D) Early stopping

Question

- 6) A discount factor  $\gamma$  close to 0 emphasizes:
- A) Immediate rewards
  - B) Long-term rewards
  - C) Exploration
  - D) Deterministic policies

- 7) The exploration–exploitation dilemma refers to:
- A) Balancing policy and value networks
  - B) Choosing between trying new actions and using known good actions
  - C) Data augmentation
  - D) Hyper-parameter tuning