Q.4	i.	Explain the use of deep neural network in reinforcement learning.	**3**	01	01	02	01
	ii.	Explain the policy gradient methods.	**7**	01	01	03	01
OR	iii.	Explain the trust region policy optimization (TRPO).	**7**	01	01	03	01
Q.5	i.	What is exploration and exploitation?	**4**	01	01	02	01
	ii.	Explain the key components and types of Multi-Agent Reinforcement Learning.	**6**	01	01	02	01
OR	iii.	Explain the challenges in transfer learning.	**6**	01	01	02	01
Q.6		Attempt any two:					
	i.	Explain the role of reinforcement learning in robotics.	**5**	01	01	05	01
	ii.	Explain the role of reinforcement learning in health care.	**5**	01	01	05	01
	iii.	Explain the role of reinforcement learning in control system.	**5**	01	01	05	01

******

Total No. of Questions: 6 *Total No. of Printed Pages:4*

Enrollment No.....

Faculty of Engineering
End Sem Examination Dec 2024
RA3EL05 Reinforcement Learning

Programme: B.Tech. Branch/Specialisation: RA

Duration: 3 Hrs. **Maximum Marks: 60**

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d. Assume suitable data if necessary. Notations and symbols have their usual meaning.

			Marks	**BL**	**PO**	**CO**	**PSO**
Q.1	i.	What is the main goal of reinforcement learning? (a) Minimize errors (b) Maximize cumulative reward (c) Optimize weights (d) Minimize loss function	**1**	01	01	01	01
	ii.	In reinforcement learning, the entity that makes decisions and takes actions is known as the- (a) Environment (b) Agent (c) Reward (d) State	**1**	01	01	01	01
	iii.	Which of the following is a key characteristic of Monte Carlo methods in reinforcement learning? (a) Learning after every step (b) Waiting until the end of an episode to update values (c) Using a neural network for function approximation (d) Optimizing policy without rewards	**1**	01	01	02	01

iv.	Temporal Difference (TD) learning combines ideas from which two reinforcement learning methods? (a) Monte Carlo methods and Q-learning (b) Monte Carlo methods and dynamic programming (c) Q-learning and supervised learning (d) Policy gradients and actor-critic methods	**1**	01	01	02	01
v.	What is the main difference between Q-learning and Deep Q-Networks (DQN)? (a) Q-learning is used for continuous action spaces, while DQN is for discrete spaces (b) DQN uses a neural network to approximate Q-values, while Q-learning typically uses a table (c) Q-learning is an on-policy method, while DQN is an off-policy method (d) DQN is used only for supervised learning tasks	**1**	01	01	03	01
vi.	In policy gradient methods, which of the following is optimized? (a) State values (b) Q-values (c) The parameters of the policy function (d) The experience replay buffer	**1**	01	01	03	01
vii.	What is the main goal of imitation learning? (a) To learn an optimal policy from trial and error (b) To imitate expert behavior by learning from demonstrations (c) To maximize rewards based on the agent's own exploration (d) To create independent learning environments for agents	**1**	01	01	04	01

viii.	Which of the following methods is often used to address non-stationarity in multi-agent reinforcement learning? (a) Using a single reward function for all agents (b) Employing centralized training with decentralized execution (c) Relying solely on experience replay (d) Training each agent individually without communication	**1**	01	01	04	01
ix.	In the context of robotics, what is the primary objective of reinforcement learning? (a) To replicate human emotions in robots (b) To enable robots to autonomously learn optimal actions to achieve specific tasks (c) To use pre-defined rules for all tasks (d) To focus on a single, static environment.	**1**	01	01	05	01
x.	Which reinforcement learning technique is often used in healthcare to manage long-term patient treatments? (a) Temporal Difference (TD) learning (b) Multi-Armed Bandit approach (c) Q-learning (d) Deep Q-Networks (DQN)	**1**	01	01	05	01
Q.2	i. What is reinforcement learning?	**2**	01	01	01	01
	ii. Define the following- (a) Value functions (b) State value function (c) Action value function	**3**	01	01	01	01
	iii. Explain the components of a reinforcement learning system.	**5**	01	01	02	01
OR	iv. Explain the Markov Decision Process (MDP).	**5**	01	01	01	01
Q.3	i. What is eligibility trace in RL?	**4**	01	01	02	01
	ii. Write the difference between Monte Carlo and Temporal Difference (TD) Learning.	**6**	01	01	02	01
OR	iii. Explain the SARSA algorithm.	**6**	01	01	01	01

Marking Scheme
RA3EL05(T) Reinforcement learning

Q.1	i)	B. Maximize cumulative reward	**1**
	ii)	B. Agent	**1**
	iii)	B. Waiting until the end of an episode to update values	**1**
	iv)	B. Monte Carlo methods and dynamic programming	**1**
	v)	B. DQN uses a neural network to approximate Q-values, while Q-learning typically uses a table.	**1**
	vi)	C. The parameters of the policy function	**1**
	vii)	B. To imitate expert behavior by learning from demonstrations	**1**
	viii)	B. Employing centralized training with decentralized execution	**1**
	ix)	B. To enable robots to autonomously learn optimal actions to achieve specific tasks	**1**
	x)	A. Temporal Difference (TD) learning	**1**
Q.2	i.	What is reinforcement learning? Definition 2 marks	**2**
	ii.	Define the following (1 mark each A. Value functions B. State value function C. Action value function	**3**
	iii.	Explain the Components of a Reinforcement Learning System. Atleast 5 components 1 marks each	**5**
OR	iv.	Explain the Markov Decision Process (MDP). Definition 1 mark Key components explanation 3 marks Formulas 1 mark	**5**
Q.3	i.	What is eligibility test? Also wrote its update rule. Definition 2 marks Update rule 2 marks	**4**
	ii.	Write the difference between monte carlo and Temporal Difference (TD) Learning. Atleast 6 difference 1 mark each	**6**
OR	iii.	Explain the SARSA Algorithm. Definition 1 mark Key components 2 marks Update rule 2 marks Uses 1mark	**6**
Q.4	i.	Explain the use of deep neural network in reinforcement learning. Atleast 3 uses 1 mark each	**3**
	ii.	Explain the policy gradient methods. Direct Policy Optimization 2 marks Objective Function 2 marks Policy Gradient Theorem 2 marks Application 1 mark	**7**
OR	iii.	Explain the trust region policy optimization (TRPO). Definition 1mark Key components 4 marks Application 2 marks	**7**
Q.5	i.	What is exploration and exploitation? exploration and exploitation 2 marks each	**4**
	ii.	Explain the key components and types of Multi-Agent Reinforcement Learning. Atleast 6 Key component 1 mark each	**6**
OR	iii.	Explain the Challenges in Transfer Learning. Definition 1mark Types and its explanation 4 marks Application 1 mark	**6**

P.T.O.

[2] [3]

Q.6 Attempt any two:

i. Explain the reinforcement learning in robotics. **5**

ii. Explain the reinforcement learning in health care. **5**

iii. Explain the reinforcement learning in control system. **5**

******