National University of Engineering (UNI)
School of Computer Science
Syllabus 2027-I

1. COURSE

AI265. Reinforcement Learning (Elective)

2. GENERAL INFORMATION

2.1 Course	:	AI265. Reinforcement Learning
2.2 Cycle	:	CS300 Cycle.
2.3 Credits	:	4
2.4 Hours	:	2 HT; 4 HP;
2.5 Duration of the period	:	16 weeks
2.6 Type of course	:	Elective
2.7 Learning modality	:	Face to face
2.8 Prerrequisites	:	AI263. Introduction to Machine Learning. (6^{th} Sem)

3. PROFESSORS

Meetings after coordination with the professor

4. INTRODUCTION TO THE COURSE

Reinforcement Learning (RL) is fundamental for autonomous systems. This course covers key algorithms such as Q-Learning, DQN, and Policy Gradients, with applications in robotics and games.

5. GOALS

- Design RL agents for stochastic environments.
- Implement RL algorithms using modern frameworks (OpenAI Gym, TensorFlow/PyTorch).
- Analyze ethical problems in autonomous systems based on RL.

6. COMPETENCES

AG-C08) Problem Analysis: Identifies, formulates, and analyzes complex computing problems. (Usage)

AG-C09) Design and Development of Solutions: Designs, implements, and evaluates solutions for complex computing problems. (Usage)

7. TOPICS

Unit 1: Foundations of RL (18 hours)	
Competences Expected: AG-C08,AG-C09	
Topics	**Learning Outcomes**
• Introduction to the Reinforcement Learning paradigm: agent, environment, actions, states, rewards. • Formalization of sequential problems: Markov Decision Processes (MDPs) and Partially Observable MDPs (POMDPs). • Concept of horizon and discount factor (gamma). • The Bellman Equation: foundation of optimality for state-value (V) and action-value (Q) functions. • Dynamic Programming Methods: Iterative Policy Evaluation and Improvement. • Monte Carlo Methods: learning from complete episodes. • Temporal Difference (TD) Algorithms: TD(0), SARSA, and the bias-variance trade-off. • Q-Learning (off-policy): derivation, convergence properties, and parameters (learning rate, epsilon-greedy).	• Model problems as MDPs [Evaluar] • Implement Q-Learning [Usar]
Readings : [SB18]	

Unit 2: Advanced Classical RL Methods (14 hours)	
Competences Expected: AG-C08,AG-C09	
Topics	**Learning Outcomes**
• Limitations of tabular methods and the curse of dimensionality. • Value Function Approximation: replacing tables with parametric models (linear, non-linear). • Temporal Difference Gradient Algorithms: minimization of the Bellman error. • Eligibility Traces: unification of Monte Carlo and TD (TD(δ)). • Integrated Planning and Learning: Environment models and Dyna-Q algorithms. • Policy-Based Methods (Policy Search): introduction to policy gradient. • Exploration vs. Exploitation Trade-off: advanced strategies (UCB, entropy-based exploration).	• Model problems as MDPs [Evaluar] • Implement Q-Learning [Usar]
Readings : [SB18]	

Unit 3: Deep Reinforcement Learning (20 hours)	
Competences Expected: AG-C08,AG-C09	
Topics	**Learning Outcomes**
• The Deep Learning revolution in RL: training stabilization. • Deep Q-Networks (DQN): architecture, loss function, and stabilization mechanisms (Target Network, Experience Replay). • DQN extensions: Double DQN, Dueling DQN, Prioritized Experience Replay. • Stochastic Policy for continuous action spaces: introduction to Actor-Critic methods. • Actor-Critic Methods: policy gradient decomposition, advantage function (A2C). • Advanced Actor-Critic Methods: A3C (Asynchronous Advantage Actor-Critic) and its parallel architecture. • Direct Policy Gradient: REINFORCE with and without baseline. • Introduction to maximum entropy methods (Soft Actor-Critic - SAC) and offline RL.	• Train DQN agents in complex environments [Evaluar] • Optimize policies with A3C [Usar]
Readings : [Mni+15], [Sil+17]	

Unit 4: Ethics, Safety, and Applications in RL (12 hours)	
Competences Expected: AG-C08,AG-C09	
Topics	**Learning Outcomes**
• Biases in reward function design: misspecification, misaligned rewards, and reward hacking. • Challenges of generalization and robustness in RL-trained agents. • Safety in autonomous systems: avoiding catastrophic failures, Safe RL, and AI Alignment. • Case studies: analysis of failures in simulations and real-world robots. • Adversarial attacks on RL: attacks on sensors, policy function, and agent's reward function. • Social implications: automation, decision-making, and transparency (interpretable agents). • Ethical frameworks for the development and deployment of autonomous agents.	• Identify ethical risks in RL applications [Familiarizarse] • Propose mitigation strategies [Usar]
Readings : [SB18]	

8. WORKPLAN

8.1 Methodology

Individual and team participation is encouraged to present their ideas, motivating them with additional points in the different stages of the course evaluation.

8.2 Theory Sessions

The theory sessions are held in master classes with activities including active learning and roleplay to allow students to internalize the concepts.

8.3 Practical Sessions

The practical sessions are held in class where a series of exercises and/or practical concepts are developed through problem solving, problem solving, specific exercises and/or in application contexts.

9. EVALUATION SYSTEM

Theory Sessions:
The theory sessions are held in master classes with activities including active learning and roleplay to allow students to internalize the concepts.

Lab Sessions:
In order to verify their competences, several activities including active learning and roleplay will be developed during lab sessions.

Oral Presentations:
Individual and team participation is encouraged to present their ideas, motivating them with additional points in the different stages of the course evaluation.

Reading:
Throughout the course different readings are provided, which are evaluated. The average of the notes in the readings is considered as the mark of a qualified practice. The use of our virtual campus allows each student to access the course information, and interact outside the classroom with the teacher and with the other students.

The final grade NF depends on the following grades:

- PE: Partial Exam.
- FE: Final Exam.
- P: Permanent Evaluation (Average of practical work and graded assignments).

The final grade NF is calculated as follows:

$$NF = \frac{PE + FE + P}{3}$$

10. BASIC BIBLIOGRAPHY

[Mni+15] Volodymyr Mnih et al. "Human-level control through deep reinforcement learning". In: *Nature* 518 (2015), pp. 529–533.

[Sil+17] David Silver et al. "Mastering the game of Go without human knowledge". In: *Nature* 550 (2017), pp. 354–359.

[SB18] Richard Sutton and Andrew Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2018.