# Assignment: Reinforcement Learning – PPO, SAC, Q-Learning

Deep Reinforcement Learning (21ARE402)

**Dead Line: 15/09/2025**

**Number of students per team: 3**

**Objective:**

To design and implement a small-scale reinforcement learning problem, experiment with different algorithms (PPO, SAC, Q-Learning/Value-based), and analyze the effect of reward shaping and algorithmic choice on performance and learning behavior.

**Instructions:**

## Part 1: Problem Statement & Literature Review (10 Marks)

1. Conduct a short literature review on reinforcement learning paradigms – Policy-based, Value-based, and Actor-Critic methods.
2. Identify the strengths and limitations of **Q-learning (Value-based)**, **PPO (Policy-gradient)**, and **SAC (Actor-Critic)**.
3. Define a **unique, simple RL environment** (grid-world, predator-prey, traffic light control, warehouse navigation, etc.) designed by your team. Clearly specify:
   - State space
   - Action space
   - Reward design (including shaped rewards)
   - Termination conditions

## Part 2: Implementation & Experimentation (20 Marks)

1. Implement your chosen environment (discrete or continuous). Keep it computationally simple.
2. Train **at least two RL algorithms** from different families (e.g., PPO vs Q-Learning, PPO vs SAC, SAC vs Q-Learning).
3. Implement a **baseline heuristic** (e.g., random policy, greedy strategy) for comparison.
4. Document hyperparameters, training steps, and compute time. Limit training to ≤ 30 minutes per run.
5. Ensure reproducibility: fix seeds, log library versions, and provide runnable code.

## Part 3: Reward Shaping & Visualization (05 Marks)

1. Apply reward shaping – design components (positive, negative, efficiency rewards).
2. Visualize and analyze the effect of shaped rewards:
   - Learning curves (average return vs. steps)
   - Policy/Value visualization (heatmaps, greedy arrows, or action histograms)
   - Reward component breakdowns

## Part 4: Analysis & Insights (05 Marks)

1. Compare the algorithms on:
   - Learning stability
   - Convergence speed
   - Robustness/generalization
   - Computational efficiency
2. Discuss how algorithm choice and reward design impacted performance.

3. Reflect on **failure cases, limitations, and potential improvements**.

**Part 5: Report and Presentation (10 Marks)**

- Prepare a detailed report (2–3 pages) summarizing:
    - Literature review
    - Problem formulation
    - Experimental setup and results
    - Reward shaping impact
    - Comparative insights
- Create a short presentation with visuals (learning curves, demo video/GIF, key findings).

## Submission Guidelines

- Submit your **report, code, and presentation slides** electronically before the deadline.
- Code must be **well-documented and reproducible** (single script/notebook preferred).
- Include a short **demo GIF/video (≤30s)** of your agent's behavior.
- Each team must propose a **unique environment/problem statement** (no duplicates allowed).

## Grading Rubric (50 Marks Total)

- Problem Statement & Literature Review: **10 marks**
- Implementation & Experimentation: **20 marks**
- Reward Shaping & Visualization: **10 marks**
- Analysis & Insights: **05 marks**
- Report and Presentation: **05 marks**

**Bonus (+5 Marks):**
Awarded for:

- Designing a **novel/custom environment** beyond standard RL benchmarks.
- Using **innovative evaluation metrics** (e.g., robustness under noise, fairness metrics, reward ablation studies).