

The Hong Kong University of Science and Technology
UG Course Syllabus (Fall 2026-27)

[Course Title] Reinforcement Learning

[Course Code] COMP4212

[No. of Credits] 3

[Any pre-/co-requisites] (COMP 2012 OR COMP 2012H) AND COMP 2211

Math: You should have familiar background in Linear Algebra (e.g., matrix inversion) and Probability (e.g., expectation, sampling).

Machine Learning: Basic machine learning knowledge (e.g., gradient backpropagation) and deep learning knowledge (e.g., neural network) as needed.

Programming: Python (necessary for assignment), PyTorch (better)

Name: CHEN, Long

Email: longchen@ust.hk

Office Hours: Appointment-based, CYT-3003, Cheng Yu Tung Building, HKUST

Course Description

[Briefly describe the course content, key topics or themes, objectives, methods of instruction, e.g., lectures, discussions, projects].

Reinforcement learning (RL) is a computational learning approach where an agent tries to maximize the total amount of reward it receives while interacting with a complex and uncertain environment. It not only shows strong performance in lots of games (such as Go), but also becomes an essential technique in many today's real-world applications (such as LLM training, and embodied AI). This course aims to teach the fundamentals and the advanced topics of RL. The course content includes the introduction of basic RL elements (including MDP, dynamic programming, policy iteration), value-based approaches (DQN), policy-based approaches (policy gradient, actor critic), model-based RL, and RL techniques in today's computer vision or AI applications. To better enhance the understanding, we will also contain some Python/Pytorch implementations.

Syllabus

- Markov Decision Processes
- Dynamic Programming
- Monte Carlo and Temporal Difference Learning
- Q-Learning
- DQN and advanced techniques
- Policy Gradient
- Actor Critic
- Advanced Policy Gradient
- Continuous Controls
- Imitation Learning
- Model-based RL

Intended Learning Outcomes (ILOs)

By the end of this course, students should be able to:

1. Being familiar with different foundational RL concepts, and differentiate RL from other machine learning algorithms.
2. Formulate real-world problems in RL framework (i.e., Markov Decision Processes (MDPs)).
3. Understand and implement value-based methods (e.g., Q-learning, Deep Q-Networks (DQN)) and policy-based methods (e.g., Policy Gradients, Proximal Policy Optimization (PPO)) to solve discrete/continuous control problems
4. Understand advanced value-based and policy-based methods, and apply these techniques into designing more advanced algorithms.
5. Evaluate RL algorithms from different metrics.
6. Know the strengths and weaknesses of state-of-the-art RL algorithms, and their applications in real-world scenarios.

Assessment and Grading

This course will be assessed using criterion-referencing and grades will not be assigned using a curve. Detailed rubrics for each assignment are provided below, outlining the criteria used for evaluation.

Assessments:

[List specific assessed tasks, exams, quizzes, their weightage, and due dates; perhaps, add a summary table as below, to precede the details for each assessment.]

Assessment Task	Contribution to Overall Course grade (%)	Due date
Assignment	40%	30/11/2026*
Midterm examination	20%	30/11/2026*
Final examination	40%	30/11/2026*

* Assessment marks for individual assessed tasks will be released within two weeks of the due date.

Mapping of Course ILOs to Assessment Tasks

Assessed Task	Mapped ILOs	Explanation
Assignment	ILO1, ILO2, ILO3. ILO4, ILO5, ILO6	This task will cover all these aspects by designing different quizzes. The coding assignment will help to students to implement and understand these new algorithms.
Midterm examination	ILO1, ILO2, ILO3. ILO4, ILO5, ILO6	The exam will cover all these aspects.
Final examination	ILO1, ILO2, ILO3. ILO4, ILO5, ILO6	The exam will cover all these aspects.

Grading Rubrics

Criterion / ILO	Grade A (Excellent)	Grade B (Good)	Grade C (Satisfactory)	Grade D (Marginal Pass)	Grade F (Fail)
1. RL Foundations & Conceptual Mastery (ILO 1)	Demonstrates a comprehensive, deep grasp of foundational concepts. Effortlessly contrasts RL with supervised/unsupervised paradigms and explains subtle nuances.	Shows solid understanding of core RL concepts. Accurately differentiates RL from other ML paradigms with minimal guidance.	Displays adequate knowledge of core RL concepts. Can differentiate RL from other ML paradigms in standard, familiar contexts.	Possesses threshold knowledge of core RL terminology, but struggles to articulate distinctions from other ML paradigms.	Demonstrates insufficient or erroneous understanding of fundamental RL concepts and paradigms.
2. Problem Formulation (MDPs) (ILO 2)	Formulates complex real-world problems into MDPs (S, A, P, R, γ) with significant creativity, optimal state representation, and well-balanced reward shaping.	Correctly formulates standard and moderately complex problems into MDPs with well-defined components and appropriate reward structures.	Successfully models basic or familiar problems into standard MDP frameworks with minor assistance or minor formulation flaws.	Constructs basic MDPs for simple tasks, but with incomplete or flawed state/action spaces and sub-optimal reward design.	Fails to correctly model problems as MDPs; unable to define state, action, transition, or reward components.
3. Implementation of Core Algorithms (ILO 3)	Implements and debugs both value-based (Q-learning, DQN) and policy-based (Policy Gradients, PPO) methods flawlessly for discrete/continuous control.	Implements value-based and policy-based algorithms accurately. Code is functional with good structure and minimal assistance needed for bug fixes.	Implements core algorithms accurately for simple or familiar environments, but struggles with hyperparameter tuning or edge cases.	Produces partially functional implementations. Struggles to implement policy-based methods or continuous control mechanisms independently.	Code is non-functional or missing. Demonstrates inability to translate mathematical algorithms (DQN, PPO) into functional code.
4. Advanced Techniques & Design (ILO 4)	Demonstrates high scholarship and creativity by designing or extending advanced algorithms (e.g., actor-critic variants, multi-agent) beyond core requirements.	Competently analyzes and applies advanced value/policy-based techniques to improve baseline model performance.	Understands key principles of advanced methods and can adapt existing algorithm templates to solve standard tasks.	Shows basic awareness of advanced methods, but lacks the ability to modify or apply them effectively to new problems.	Shows no understanding of advanced RL methods; unable to apply or reason about algorithm modifications.

		on non-trivial tasks.			
5. Evaluation & Empirical Analysis (ILO 5)	Conducts rigorous evaluation using multi-metric analysis (sample efficiency, variance, stability, wall-clock time) with insightful interpretation.	Evaluates models using standard metrics (mean reward, convergence rates) and accurately interprets training plots and performance trade-offs.	Evaluates algorithms using basic reward curves. Able to draw correct conclusions about model convergence on familiar tasks.	Uses minimal metrics (e.g., raw return only). Exhibits weak critical analysis when interpreting training dynamics or failures.	Lacks necessary evaluation skills; fails to present valid metrics, plots, or performance analyses.
6. Real-World Trade-Offs & Application (ILO 6)	Critically evaluates SOTA RL strengths/weaknesses. Proposes innovative real-world applications with detailed risk and feasibility assessments.	Effectively analyzes SOTA algorithm trade-offs and correctly identifies suitable real-world applications for different RL techniques.	Recognizes basic strengths and weaknesses of popular SOTA algorithms in standard application domains.	Has threshold knowledge of where SOTA algorithms apply, but struggles to justify algorithm choice for specific scenarios.	Minimal understanding of SOTA algorithms; cannot match appropriate RL approaches to real-world problems.

Final Grade Descriptors:

[As appropriate to the course and aligned with university standards]

Grades	Short Description	Elaboration on subject grading description
A	Excellent Performance	Demonstrates a comprehensive grasp of subject matter, expertise in problem-solving, and significant creativity in thinking. Exhibits a high capacity for scholarship and collaboration, going beyond core requirements to achieve learning goals.
B	Good Performance	Shows good knowledge and understanding of the main subject matter, competence in problem-solving, and the ability to analyze and evaluate issues. Displays high motivation to learn and the ability to work effectively with others.
C	Satisfactory Performance	Possesses adequate knowledge of core subject matter, competence in dealing with familiar problems, and some capacity for analysis and critical thinking. Shows persistence and effort to achieve broadly defined learning goals.
D	Marginal Pass	Has threshold knowledge of core subject matter, potential to achieve key professional skills, and the ability to make basic judgments. Benefits from the course and has the potential to develop in the discipline.
F	Fail	Demonstrates insufficient understanding of the subject matter and lacks the necessary problem-solving skills. Shows limited ability to think critically or analytically and exhibits minimal effort towards achieving learning goals. Does not

		meet the threshold requirements for professional practice or development in the discipline.
--	--	---

Course AI Policy

The students cannot use AI tools in all these assignment tasks.

Communication and Feedback

Assessment marks for individual assessed tasks will be communicated via Canvas within two weeks of submission. Feedback on assignments will include [specific details, e.g., strengths, areas for improvement]. Students who have further questions about the feedback including marks should consult the instructor within five working days after the feedback is received.

Resubmission Policy

No resubmission

Required Texts and Materials

Richard S. Sutton. Reinforcement Learning: An Introduction. Second Edition.
Kevin P. Murphy. Reinforcement Learning: An Overview.

Academic Integrity

Students are expected to adhere to the university's academic integrity policy. Students are expected to uphold HKUST's Academic Honor Code and to maintain the highest standards of academic integrity. The University has zero tolerance of academic misconduct. Please refer to [Academic Integrity | HKUST – Academic Registry](#) for the University's definition of plagiarism and ways to avoid cheating and plagiarism.

[Optional] Additional Resources

NIL