COMPSCI 687

Reinforcement Learning

Fall 2026 Syllabus

People

Instructor	Professor Philip S. Thomas (pthomas@umass.edu)
Teaching Assistants	Norman Zhang (renhaozhang@umass.edu) Shauna Choi (seohyunchoi@umass.edu) Shruti Chanumolu (schanumolu@umass.edu)

Office Hours

Day	Time	Location and Staff
Monday	10–12	CSL E211, Shruti Chanumolu
Tuesday	10–12	CSL E211, Shauna Choi
Wednesday	2–4	CSL E213, Norman Zhang

Office hours will be in-person only (not hybrid). In rare cases TAs may move their office hours to be online only (still not hybrid).

If you have a conflict with all of the TA office hour times, please send one email to all three TAs to request an additional meeting time, listing your availability over the next week. If you have questions about course material that the TAs are unable to answer, please ask Prof. Thomas the question in lecture or ask the TAs to convey the question to Prof. Thomas as one to consider answering during lecture. Note that Prof. Thomas is not available to answer questions that are not about course material (e.g., regarding personal projects, research projects, or other guidance).

Course Overview

COMPSCI 687 is a 3-credit, in-person course with three instructional contact hours per week. It meets from 8:30–9:45 a.m. on Mondays and Wednesdays in Computer Science Laboratories E110. Note that UMass course policies indicate 2 hours of work outside of lecture for every hour of lecture. With 2.5 hours of lecture per week, this equates to 5 hours of work outside of class each week.

Prerequisites and Expected Background

This course has no formally enforced prerequisite courses. It assumes a strong background in calculus, linear algebra, and strategies for proving theorems; strong programming ability; and background in machine learning comparable to COMPSCI 589 or COMPSCI 689 and artificial intelligence comparable to COMPSCI 683. The course is intended primarily for graduate Computer Science students.

Course Description

This course provides an introduction to reinforcement learning. Reinforcement learning algorithms repeatedly answer the question "What should be done next?", and they can learn via trial-and-error to answer these questions even when there is no supervisor telling the algorithm what the correct answer would have been. Applications of reinforcement learning span across medicine (How much insulin should be injected next? What drug should be given next?), marketing (What ad should be shown next?), robotics (How much power should be given to the motor?), game playing (What move should be made next?), environmental applications (Which countermeasure for an invasive species should be deployed next?), and dialogue systems (What type of sentence should be spoken next?), among many others. Broad topics covered in this course will include: Markov decision processes, reinforcement learning algorithms (model-based/model-free, batch/online, value function-based, actor-critics, policy gradient methods, etc.), and representations for reinforcement learning. Special topics may include ensuring the safety of reinforcement learning algorithms, hierarchical reinforcement learning, theoretical reinforcement learning, multi-agent reinforcement learning, fine-tuning of large language models, and connections to animal learning.

Student Learning Objectives

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

- formulate reinforcement learning problems as (partially observable) Markov decision processes and apply standard algorithms to solve them;
- explain the main ideas behind foundational reinforcement learning algorithms and concepts;
- distinguish among broad classes of reinforcement learning methods and describe how reinforcement learning methods are used to fine-tune large language models.

Course Materials

- **Course Webpage:** https://people.cs.umass.edu/~pthomas/courses/CMPSCI_687_Fall2026.html
- **Book:** None. Typed notes will be provided on the course website.
- **Gradescope:** We will use Gradescope for assignments (submission, grading, and regrade requests). You will receive an invitation to Gradescope during the first week or two of classes.
- **Canvas:** The syllabus and links to the course webpage and Gradescope will be posted in Canvas. Canvas will not otherwise be used for course content.

Course announcements will be sent to students' UMass email addresses. We will not use Piazza or a similar system. Lecture recordings will not be available. Practice tests will not be provided.

Tentative Schedule

The Fall 2026 semester provides 26 scheduled meetings for this Monday/Wednesday course. The schedule below is tentative. Topics may shift as the semester progresses; any changes will be reflected on the course webpage.

Test 1: October 14, 2026 from 7pm–9pm, room Integrative Learning Center (ILC) Room N151.

Test 2: Date to be determined, around November 23, 2026.

Final Exam: December 21, 2026 from 8am–12pm in CSL E110.

Week and Meeting Dates	Planned Topics
Week 1: September 9	Course introduction; notation; what reinforcement learning is
Week 2: September 14 and 16	Evaluative feedback and sequential decision making; 687-Gridworld; Markov decision process components, policies, trajectories, returns, and objectives

Week and Meeting Dates	Planned Topics
Week 3: September 21 and 23	Creating Markov decision processes; planning versus reinforcement learning; additional terminology and assumptions; example domains
Week 4: September 28 and 30	Black-box optimization for policy search; policy parameterizations; evaluating reinforcement learning algorithms; introduction to value functions
Week 5: October 5 and 7	State-value and action-value functions; Bellman equations; optimal value functions and Bellman optimality equations
Week 6: October 14	Policy evaluation and the beginning of policy improvement (no class October 12)
Week 7: October 19 and 21	Policy improvement and policy iteration; value iteration; the Bellman operator
Week 8: October 26 and 28	Convergence of value iteration; Monte Carlo policy evaluation
Week 9: November 2 and 4	Convergence of Monte Carlo methods; gradient-based Monte Carlo updates; temporal-difference learning
Week 10: November 9	Convergence of temporal-difference learning; function approximation; maximum-likelihood MDP models versus temporal-difference learning (no class November 11)
Week 11: November 16 and 18	Sarsa and Q-learning, including on-policy and off-policy control and convergence considerations
Week 12: November 23 and 24	Off-policy evaluation; importance sampling; high-confidence off-policy evaluation; high-confidence policy improvement (November 24 follows a Wednesday schedule)
Week 13: November 30 and December 2	TD(λ), λ-returns, eligibility traces, true online TD(λ), Sarsa(λ), and Q(λ)
Week 14: December 7 and 9	Policy-gradient methods; the policy-gradient theorem; REINFORCE; natural gradient; selected topics including hierarchical reinforcement learning, experience replay, multi-agent reinforcement learning, reinforcement learning theory, and deep reinforcement learning; introduction to fine-tuning large language models
Week 15: December 14	Reinforcement learning methods for fine-tuning large language models; course review
Final examination period: December 17–23	Cumulative final examination; the exact date and time will be announced when assigned by the University

Requirements and Grading

Attendance is required. Participation is not. Attendance will not be taken, although there will be pop-quizzes at the **start** of class on random unannounced days. If you are unable to reliably be on-time for lectures, you should not take this course.

Your grade will be based on home work assignments (≥ 5), pop quizzes (≥ 5), and tests (3). Below are the weights assigned to each:

- **Homework Assignments:** 10%
- **Pop Quizzes:** 20%
- **Tests:** 70%

Homework assignments. Homework assignments are not equally weighted. The number of points in each assignment determines its weight, with a target of roughly 100 points for a typical assignment.

Pop Quizzes. Pop quizzes will be equally weighted. They will be given at the start of lecture on randomly selected days. Each question will be marked either right or wrong, with no partial credit for nearly correct answers.

Tests. There will be three tests. They will all be closed-notes and the use of electronic devices will be prohibited. The first two will not be cumulative and will be worth 20% of your total grade each, while the third (a final exam) will be cumulative and worth 30% of your total grade. The tests will cover both material from lectures **and** material covered on homework assignments.

Grading Scale

The following table will be used for converting numerical grades at the end of the course to letter grades. Minimum values are inclusive, while maximum values are exclusive (except for 100). For example, 93 corresponds to an A, not an A-.

Minimum	Maximum	Letter
0	70	F
70	77	C
77	80	C+
80	83	B-
83	87	B
87	90	B+
90	93	A-
93	100	A

Other Policies

Late or Make-Up Work Policies

- **Late Homework Assignment Policy:** I trust your judgment when determining whether you need an extension on an assignment. You may take a 7 day extension on any and every assignment with no penalty. Assignments submitted more than 7 days late will not be accepted unless there are extenuating circumstances.
- **Regrade Request Policy:** Regrade requests for homework assignments will only be considered if they are submitted within 1 week of the grade for the assignment being announced.
- **Missed Pop-Quiz Policy:** If you are going to miss the start of lecture, e-mail Prof. Thomas (pthomas@umass.edu) prior to the start of lecture explaining your absence. If you are unable to email prior to the start of lecture, email after lecture and explain both the absence **and the reason why you were unable to message prior to lecture**. Prof. Thomas will respond within one week with a determination or to request additional evidence regarding your absence.
- **Make-Up Exam Policy:** If you have a conflict with a test or exam, email Prof. Thomas (pthomas@umass.edu) at least one week prior to the test or exam. If you are unable to attend an exam due to extenuating circumstances, email Prof. Thomas as soon as possible, explaining the reason you missed the test or exam. You will likely be asked to provide supporting evidence.

Remote Office Hours

Office hours will be held in-person. They will not be virtual (e.g., on Zoom).

Emailing Questions

Do not email questions to the TAs – you are expected to attend office hours and ask there, or to ask during lecture. If you attend office hours and and a TA is unable to answer your question, you may ask them to convey the question to Prof. Thomas. However, Prof. Thomas will not be answering email questions directly from students.

Required Syllabus Statements

University policies regarding accommodations, academic integrity, and Title IX apply to all courses. Because Prof. Thomas is a responsible employee, the applicable required statements are available on the Faculty Senate responsible-employee required syllabus statements page.

Collaboration and Generative AI Usage

You may collaborate with other students and may use generative AI on homework assignments without any restriction and without any required disclosure. Pop quizzes, tests, and exams are closed-book and closed-notes, must be completed individually without any collaboration, and without the use of any electronic devices.

Course-Specific Academic Integrity Policy

If an academic-integrity infraction is established through the University's Academic Integrity procedures, the course sanction will normally be zero credit for all graded work affected by the infraction. In addition, the final course grade will not exceed B. The University's procedures, including its notice and appeal provisions, will be followed.

Incompletes

A grade of INC may be considered when circumstances prevent a student from completing the course. Normally, an INC will be granted only if the student is passing the course at the time of the request. If an INC is granted, a written record will be filed with the departmental office specifying the percentage of work completed, the grade earned on the completed work, the work remaining, the method for completing it, and the completion deadline. The remaining work must be completed within one calendar year of enrollment in the course; otherwise, the INC becomes an IF. An earlier deadline may be established in the written agreement.

Satisfactory (SAT) Grading

Students may request SAT grading for this course. Requests must be made no later than the date in the UMass Academic Calendar marked as the *Last day to Drop with "DR"-Graduate*. Requests made after that date will not be approved. SAT corresponds to an underlying letter grade of C or higher; otherwise the recorded grade is F. Conditional SAT grading (for example, SAT only if the final grade would be below a specified threshold), is not allowed for this course.

Pronouns Statement

My pronouns are **he/him**.

Land Acknowledgement

I will not be including the standard land acknowledgment statement. In my view, as long as the university bears the name of Jeffrey Amherst—a figure associated with advocating and attempting to carry out genocide against Native peoples—such a statement would be disingenuous. For historical context, see: https://people.umass.edu/derrico/amherst/lord_jeff.html