

# University of Alberta

## CMPUT 365: Introduction to Reinforcement Learning

### LEC B1

### Winter 2024

**Instructor:** Scott Jordan

**TAs:** Shibhansh Dohare, Esraa Elelimy, Gábor Mihucz, Blanca Miller, and David Szepesvari

**Office:** ATH 3-21

**E-mail:** [sjordan@ualberta.ca](mailto:sjordan@ualberta.ca)

**Web Page:** <https://scottjordan.github.io/courses/CMPUT365-W24/>

**Office hours:** Scott: Monday 14:00 - 16:00 in ATH 3-31 (Athabasca Hall)

Blanca: Tuesday 14:00 - 16:00 in CAB 313

Gábor: Wednesday 09:00 - 11:00 in CAB 313

David: Wednesday 14:00 - 16:00 in CSC 3-50

Esra: Thursday 11:00 - 13:00 in CSC 3-26

Shibhansh: Friday 09:00 - 11:00 in CAB 3-13

**TA email address:** [cmput365@ualberta.ca](mailto:cmput365@ualberta.ca)

Do not personally email the TAs. They will only respond via [cmput365@ualberta.ca](mailto:cmput365@ualberta.ca).

**Lecture room & time:** CCIS 1-160, MWF 13:00 - 13:50

Attendance isn't mandatory although strongly encouraged.

---

## COURSE CONTENT

**Course Description:** This course provides an introduction to reinforcement learning, which focuses on the study and design of agents that interact with a complex, uncertain world to achieve a goal. We will emphasize agents that can make near-optimal decisions in a timely manner with incomplete information and limited computational resources. The course will cover Markov decision processes, reinforcement learning, planning, and function approximation (online supervised learning).

The course will be a flipped classroom, which, as [Wikipedia](#) says, is a “learner-centered model where students are exposed to the subject matter outside the classroom so that they can have more meaningful learning opportunities in the classroom”. The course will use the MOOC on Reinforcement Learning, created by faculty members in the department of computing science of the University of Alberta. Much of the lecture material and assignments will come from the MOOC. In-class time will be largely spent on discussion and thinking about the material, with some supplementary lectures. The idea is that instead of the standard model where you watch lectures in the classroom and study at home, you will watch lectures at home and “study” in the classroom. This model allows for more active learning instead of having students passively watching lectures in the classroom.

**Course Prerequisites:** CMPUT 175 or CMPUT 275, CMPUT 267 or permission from the instructor.

The course will use Python 3. We will use elementary ideas of probability, calculus, and linear algebra, such as expectations of random variables, conditional expectations, partial derivatives, vectors and matrices. Students should either be familiar with these topics or be ready to pick them up quickly as needed by consulting outside resources.

**Course Objectives and Expected Learning Outcomes:** By the end of the course, you will have a solid grasp of the main ideas in reinforcement learning, which is the primary approach to statistical decision-making. Any student who understands the material in this course will understand the foundations of much of modern probabilistic artificial intelligence (AI) and be prepared to take more advanced courses or to apply AI tools and ideas to real-world problems. That person will be able to apply these tools and ideas in novel situations -- e.g., to determine whether the methods apply to this situation and, if so, which will work most effectively. They will also be able to assess claims made by others, with respect to both software products and general frameworks, and also be able to appreciate some new research results.

With a focus on AI as the design of agents learning from experience to predict and control their environment, topics will include:

- Markov decision processes,
- planning by approximate dynamic programming,
- Monte Carlo and temporal difference learning for prediction,
- Monte Carlo, Sarsa and Q-learning for control,
- Dyna and planning with a learned model,
- prediction and control with function approximation.

---

## LEARNING RESOURCES

**Required Textbook and/or Other Major Course Materials:** All course reading material will be available online. We will be using videos from the Reinforcement Learning Specialization on Coursera, which was created by faculty members in the department of computing science of the University of Alberta.

We will be using the following textbook extensively:

*Sutton and Barto, Reinforcement Learning: An Introduction, MIT Press. 2nd Edition.*

The book is available from the bookstore or online as a pdf in the link below:

<http://www.incompleteideas.net/book/the-book-2nd.html>

**On-Line Homework Disclaimer:** On-line homework is a component of this course and is provided by a third-party company, Coursera, through a partnership with the University of Alberta. University of Alberta students can get free access to the online content of the Reinforcement Learning Specialisation via the steps outlined [here](#). You are required to set up a free coursera account with your ualberta.ca email address. You will be automatically enrolled in a private version of the course for this class.

**Academic Success Centre:** The [Academic Success Centre](#) provides professional academic support to help students strengthen their academic skills and achieve their academic goals. Individual advising, appointments, and group workshops are available year round in the areas of Accessibility, Communication, Learning, and Writing Resources. Modest fees apply for some services.

---

### GRADE EVALUATION

Assessment	Weight	Date
Quizzes (80% pass)	20%	Various
Assessments (notebooks on Coursera)	30%	Various
Midterm exam	20 %	Feb 12, 2024
Final exam	30%	Apr 25, 2024*

\* Students must verify this date on BearTracks when the Final Exam Schedule is posted.  
Grades are unofficial until approved by the Department and/or Faculty offering the course.

**Statement of Expectations for AI Use:** The primary goal of this course is to foster *individual* critical, creative thinking, and problem-solving skills related to reinforcement learning and, more broadly, machine learning. Thus, in order to achieve such learning outcomes, students can submit each practice quiz and graded assignment multiple times, which allows for many learning opportunities. Therefore, the use of advanced AI-tools based on large-language models such as ChatGPT or Bard is strictly prohibited for all quizzes and graded assignments. The only exception is their use for Python-related queries (but the use of such tools to help with the programming assignments themselves is still strictly prohibited). As stated in the university's [AI-Squared - Artificial Intelligence and Academic Integrity](#) webpage, "learning is not only about the product; learning is also about the process of acquiring new knowledge or learning ways to think and reason."

IMPORTANT: Failure to abide by this guideline may be considered an act of cheating and a violation as outlined in the relevant sections of University of Alberta [Code of Student Behaviour](#).

**Re-evaluation of midterm exams:** Students will have access to their midterm exam during an exam viewing period. A student who has concerns about how specific questions of their midterm exam were marked can submit a request to the instructor, via email, within two weeks the date they received their marked exam. The request should specify (1) which question is to be re-evaluated, (2) the rationale for such a request, and (3) the proposed marks. Importantly, once a request for re-evaluation is submitted, it is up to the instructor's discretion to adjust the marks. *Students won't be allowed to take their midterm exams with them, nor to take pictures of them*, so in case of concerns the student is advised to take notes during the exam viewing period. The TAs are not authorized to weigh in on the midterm exams, this is something only the instructor can do. *Notice marks can also go down once a question is re-evaluated.*

**Past or Representative Evaluative Material:** The quizzes and graded assignments are an excellent example of representative material, as well as the exercises in the course's textbook. Additional representative material will be discussed in the classroom and later posted on eClass.

**Format of Exams:** The midterm exams will be held in a room still to be defined, from 13:00 - 13:50 (50 minutes). The date, time, and place of the final exam is scheduled by the office of registrar; the final will be a 2-hour exam. *The exams will be closed book and written in real-time.*

**Format of Quizzes and Assignments:** You are allowed to consult the textbook and materials on eClass while solving the assignments. You are also allowed to discuss the assignments with your classmates. Note, however, that you are not allowed to exchange any written text, code, or to give and/or receive detailed step-by-step instructions on how to solve the proposed problems. *Moreover, as stated above, you are not allowed to use generative AI in your coursework.*

**Grade Evaluation:** At the end of the term your percentage grade will be converted into a letter grade by following the table below, extracted from the [Department Course Policies](#).

Letter	Descriptor	Interpretation
A+ A A-	Excellent	Consistently original thinking that extends the material, demonstrated depth and breadth in the material, ability to integrate material with other subjects, ability to analyze and synthesize material at various levels of abstraction.
B+ B B-	Good	Like an A, but not consistent over time, or weak in a specific area.
C+ C C-	Satisfactory	Understand the core material but not its subtleties, can apply it to simple situations on own and to more complex situations with hints, evidence that the material has changed the way of thinking.
D+	Poor	Understand some of the core material but not its subtleties, can apply it to simple situations but often needs assistance, evidence that the material has had some change on the way of thinking.
D	Minimal Pass	Shows some understanding of parts of the material, cannot apply it without some direction, little evidence that the material has changed the way of thinking.
F	Failure	Little evidence of understanding of even the surface issues, poor analysis and synthesis, inability to apply the material.

**Grade adjustment:** In total, grades add up to 100% and students are being allowed to not submit 2 quizzes and 1 graded assignments (see below).

#### **Exam Conduct:**

- Your student photo I.D. is required at exams to verify your identity.
- Students will not be allowed to begin an examination after it has been in progress for 30 minutes. Students must remain in the exam room until at least 30 minutes has elapsed.
- All cell phones must be turned off and stored in your bags.

---

## **POLICIES FOR LATE AND MISSED WORK**

**Late Policies:** Late submissions will not be accepted but students are allowed to miss some submissions, as outlined below.

**Missed quiz:** There are a total of 12 weekly quizzes and students are expected to do all of them. But, due to the fact that issues sometimes arise, students have to complete 10 of the 12 quizzes to get the full 12% participation mark. All the quizzes will be due at 23:59:59 of the day mentioned in

the schedule. Each quiz has equal weight (1.5%). *Please note students can miss 2 instances of the quiz without penalty.*

**Missed graded assignment:** There are 8 graded assignments in the form of a python notebook. Students should do all of them but, due to the fact that issues sometimes arise, students' final marks will be the sum of their top 7 assignment submissions. Effectively, students only need to complete 7 of the 8 assignments to get the full 30% marks. All the assignments will be due at 23:59:59 of the day mentioned in the schedule. Each graded assignment has equal weight. *Please note students can miss 2 instances of the quiz without penalty.*

**Missed midterm exam:** A student who cannot complete a midterm exam due to incapacitating illness, severe domestic affliction or other compelling reasons can apply for an excused absence. To apply for an excused absence, you must contact the instructor within two working days of missing the assessment or as soon as possible. If an excused absence is granted, then the weight of the missed midterm will be deferred to the final examination. An excused absence is a privilege and not a right. There is no guarantee that an absence will be excused. Misrepresentation of facts to gain an excused absence is a serious breach of the Code of Student Behaviour.

**Deferred Final Examination:** A student who cannot write the final examination due to incapacitating illness, severe domestic affliction or other compelling reasons can apply for a deferred final examination. Such an application must be made to the student's Faculty office within two working days of the missed examination and must be supported by appropriate documentation or a Statutory Declaration (see calendar on [Attendance](#)). Deferred examinations are a privilege and not a right; there is no guarantee that a deferred examination will be granted. The Faculty may deny deferral requests in cases where less than 50% of term work has been completed. Misrepresentation of facts to gain a deferred examination is a serious breach of the Code of Student Behaviour.

**Missed Term Work or Final Exam Due to Non-medical Protected Grounds (e.g., religious beliefs):** When a term assessment or final exam presents a conflict based on [non-medical protected grounds](#), students must apply to the Academic Success Centre for accommodations via their [Register for Accommodations website](#). Students can review their eligibility and choose the application process specific for *Accommodations Based on Non-medical Protected Grounds*.

It is imperative that students review the dates of all course assessments upon receipt of the course syllabus, and apply *as soon as possible* to ensure the timely application of the accommodation. Students who apply later in the term may experience unavoidable delays in the processing of the application, which can affect the accommodation.

---

## REMOTE DELIVERY CONSIDERATIONS

**Technology for Remote Learning:** To successfully participate in remote learning in this course, it is recommended that students have access to a computer with an internet connection that can support the tools and technologies the University uses to deliver content, engage with instructors, TAs, and fellow students, and facilitate assessment and examinations. Please refer to [Technology for Remote Learning - For Students](#) for details. If you encounter difficulty meeting the technology recommendations, please email the Dean of Students Office (dosdean@ualberta.ca) directly to explore options and support.

Please contact the instructor by the add/drop deadline September 18 if you do not have access to the minimum technology recommended. The instructor will make arrangements for accommodating students who contact the instructor before this date. Failure to do so may result in a zero in any assessment that depends on the minimum technology.

**Student Resources for Remote Learning:** Online learning may be new to you. Check out tips for success and find out more about online learning on the [Campus Life](#) page, and specifically on the [Academic Skills Online & Remote Delivery Resources](#) page.

---

## STUDENT RESPONSIBILITIES

**Academic Integrity and Student Conduct:** The University of Alberta is committed to the highest standards of academic integrity and honesty, as well as maintaining a learning environment that fosters the safety, security, and the inherent dignity of each member of the community, ensuring students conduct themselves accordingly. Students are expected to be familiar with the standards of academic honesty and appropriate student conduct, and to uphold the policies of the University in this respect. Students are particularly urged to familiarize themselves with the provisions of the [Code of Student Behaviour](#) and the [Student Conduct Policy](#), and avoid any behaviour that could potentially result in suspicions of academic misconduct (e.g., cheating, plagiarism, misrepresentation of facts) and non-academic misconduct (e.g., discrimination, harassment, physical assault). Academic and non-academic misconduct are taken very seriously and can result in suspension or expulsion from the University.

All students are expected to consult the [Academic Integrity website](#) for clarification on the various academic offences. All forms of academic dishonesty are unacceptable at the University. Any suspected academic offence in this course will be reported to the College of Natural and Applied Sciences. Suspected cases of non-academic misconduct will be reported to the Dean of Students. The College, the Faculty of Science, and the Dean of Students are committed to student rights and

responsibilities, and adhere to due process and administrative fairness, as outlined in the Code of Student Behaviour and the Student Conduct Policy. Anyone who is found in violation is likely to receive a sanction. Typical sanctions for academic misconduct include conduct probation, a mark reduction or a mark of 0 on an assessment, a grade reduction or a grade of F in a course, a remark on the transcript, and a recommendation for suspension or expulsion. Sanctions for non-academic misconduct include conduct conditions, fines, suspension of essential or non-essential University services and resources, and suspension or expulsion from the University.

**Appropriate Collaboration:** You are also allowed to discuss the quizzes and assignments with your classmates. Note, however, that you are not allowed to exchange any written text, code, or to give and/or receive detailed step-by-step instructions on how to solve the proposed problems. We refer students to the University of Alberta's webpage on [How to Avoid Inappropriate Collaboration](#).

**Cell Phones:** Cell phones are to be turned off during lectures, labs and seminars.

**Students Eligible for Accessibility-Related Accommodations:** In accordance with the University of Alberta's [Discrimination, Harassment, and Duty to Accommodate policy](#), accommodation support is available to eligible students who encounter limitations or restrictions to their ability to perform the daily activities necessary to pursue studies at a post-secondary level due to medical conditions and/or non-medical protected grounds. Accommodations are coordinated through the [Academic Success Centre](#), and students can learn more about eligibility on the [Register for Accommodations website](#).

**Recording and/or Distribution of Course Materials:** Audio or video recording, digital or otherwise, of lectures, labs, seminars or any other teaching environment by students is allowed only with the prior written consent of the instructor or as a part of an approved accommodation plan. Student or instructor content, digital or otherwise, created and/or used within the context of the course is to be used solely for personal study, and is not to be used or distributed for any other purpose without prior written consent from the content author(s).

### **Learning and Working Environment:**

The Faculty of Science is committed to ensuring that all students, faculty and staff are able to work and study in an environment that is safe and free from discrimination, harassment, and violence of any kind. It does not tolerate behaviour that undermines that environment. This includes virtual environments and platforms.

If you are experiencing harassment, discrimination, fraud, theft or any other issue and would like to get confidential advice, please contact any of these campus services:

- [Office of Safe Disclosure & Human Rights](#): *A safe, neutral and confidential space to disclose concerns about how the University of Alberta policies, procedures or ethical standards are being applied. They provide strategic advice and referral on matters such as discrimination, harassment, duty to accommodate and wrong-doings. Disclosures can be made in person or online using the [Online Reporting Tool](#).*



- [University of Alberta Protective Services](#): Peace officers dedicated to ensuring the safety and security of U of A campuses and community. Staff or students can contact UAPS to make a report if they feel unsafe, threatened, or targeted on campus or by another member of the university community.
- [Office of the Student Ombuds](#): A confidential and free service that strives to ensure that university processes related to students operate as fairly as possible. They offer information, advice, and support to students, faculty, and staff as they deal with academic, discipline, interpersonal, and financial issues related to student programs.
- [Office of the Dean of Students](#): They can assist students in navigating services to ensure they receive appropriate and timely resources. For students who are unsure of the support they may need, are concerned about how to access services on campus, or feel like they may need interim support while you wait to access a service, the Dean of Students office is here to help.

**Feeling Stressed, Anxious, or Upset?** It's normal for us to have different mental health experiences throughout the year. Know that there are people who want to help. You can reach out to your friends and access a variety of supports available on and off campus at the [Need Help Now](#) webpage or by calling the 24-hour Distress Line: 780-482-4357 (HELP).

**Student Self-Care Guide:** This [Self-Care Guide](#), originally designed by the Faculty of Native Studies, has broader application for use during students' learning. It provides some ideas and strategies to consider that can help navigate emotionally challenging or triggering material.

---

Policy about course outlines can be found in [Course Requirements, Evaluations Procedures and Grading](#) of the University Calendar.

**Land Acknowledgement:** The University of Alberta respectfully acknowledges that we are situated on Treaty 6 territory, traditional lands of First Nations and Métis people.

To learn more about the significance of this land acknowledgement, please read [this](#) useful article and associated links to more information.

**Disclaimer:** Any typographical errors in this Course Outline are subject to change and will be announced in class. The date of the final examination is set by the Registrar and takes precedence over the final examination date reported in this syllabus.

This syllabus was originally generated by Marlos Machado and has been reused with permission. Copyright of the original material belongs to him.

**Copyright:** Dr. Marlos Cholodovskis Machado, Department of Computing Science, Faculty of Science, University of Alberta (2023).

Copyright of new material

**Copyright:** Dr. Scott Michael Jordan, Department of Computing Science, Faculty of Science,  
University of Alberta (2024).

---

## Course Schedule & Assigned Readings

Week	Date	Topic	Deadlines (all due at 12:59:59)	Readings
1	Mon, Jan 8	Course overview Discussion about what is reinforcement learning		
1	Wed, Jan 10	Background review: Probability, statistics, linear algebra, and calculus		
1	Fri, Jan 12	Fundamentals of RL: An introduction to sequential decision-making	Quiz: Sequential decision-making	Chapter 2, up to §2.7 (pp. 25-36), and §2.10 (pp. 42-44)
2	Mon, Jan 15	Fundamentals of RL: An introduction to sequential decision-making		
2	Wed, Jan 17	Fundamentals of RL: An introduction to sequential decision-making	Program. assignment (Bandits & exploration / exploitation)	
2	Fri, Jan 19	Fundamentals of RL: Markov decision processes (MDPs)	Quiz: MDPs	Chapter 3, up to §3.3 (pp. 47-56)
3	Mon, Jan 22	Fundamentals of RL: Markov decision processes (MDPs)		
3	Wed, Jan 24	Fundamentals of RL: Markov decision processes (MDPs)		
3	Fri, Jan 26	Fundamentals of RL: Value functions & Bellman equations	Quiz: Value functions & Bellman equations 1	Chapter 3, §3.5-§3.8 (pp. 58-69)
4	Mon, Jan 29	Fundamentals of RL: Value functions & Bellman equations		
4	Wed, Jan 31	Fundamentals of RL: Value functions &	Quiz: Value functions & Bellman	

		Bellman equations	equations 2	
4	Fri, Feb 2	Fundamentals of RL: Dynamic programming	Quiz: Dynamic programming	Chapter 4, §4.1-§4.4 (pp. 73-84); §4.6-§4.7 (pp. 86-89)
5	Mon, Feb 5	Fundamentals of RL: Dynamic programming		
5	Wed, Feb 7	Fundamentals of RL: Dynamic programming	Program. Assignment (Optimal policies with dynamic programming)	
5	Fri, Feb 9	Midterm Review		
6	Mon, Feb 12	<b>Midterm Exam</b>		
6	Wed, Feb 14	Sample-based learning methods: MC methods for Prediction & Control		Chapter 5, up to §5.5 (pp. 91-108); §5.10 (pp. 115-116)
6	Fri, Feb 16	Sample-based learning methods: MC methods for Prediction & Control	Quiz: Off-policy Monte Carlo	
	Feb 19-22	Reading Week		
7	Mon, Feb 26	Sample-based learning methods: MC methods for Prediction & Control		
7	Wed, Feb 28	Sample-based learning methods: TD learning for prediction	Quiz: Advantages of TD	Chapter 6, up to §6.3 (pp. 119-128)
7	Fri, Mar 1	Sample-based learning methods: TD learning for prediction		
8	Mon, Mar 4	Sample-based learning methods: TD learning for prediction	Program. Assignment (Policy evaluation with TD learning)	
8	Wed, Mar 6	Sample-based learning methods: TD learning for control	Quiz: Expected Sarsa	Chapter 6, §6.4-§6.6 (pp. 129-134); §6.10 (p. 138)

8	Fri, Mar 8	Sample-based learning methods: TD learning for control		
9	Mon, Mar 11	Sample-based learning methods: TD learning for control	Program. assignment (Q-learning & Expected Sarsa)	
9	Wed, Mar 13	Prediction and Control with FA: On-policy prediction with approx.	Quiz: On-policy prediction with approximation	Chapter 9, up to §9.4 (pp. 197-209)
9	Fri, Mar 15	Prediction and Control with FA: On-policy prediction with approx.		
10	Mon, Mar 18	Prediction and Control with FA: On-policy prediction with approx.	Program. assignment (Semi-gradient TD(0) with state aggregation)	
10	Wed, Mar 20	Prediction and Control with FA: Constructing features for prediction	Quiz: Constructing Features for Prediction	Chapter 9, §9.5 (pp. 210-222) and §9.12 (pp. 236-237)
10	Fri, Mar 22	Prediction and Control with FA: Constructing features for prediction		
11	Mon, Mar 25	Prediction and Control with FA: Constructing features for prediction	Program. assignment (Semi-gradient TD with a neural network)	
11	Wed, Mar 27	Prediction and Control with FA: Control with approximation	Quiz: Control with approximation	Chapter 10, up to §10.1 (pp. 243-248); §10.3 (pp. 249-252), §10.6 (p. 256)
11	Fri, Mar 29	Good Friday (No Class)		
12	Mon, Apr 1	Easter Monday (No Class)		
12	Wed, Apr 3	Prediction and Control with FA: Control with approximation		
12	Fri, Apr 5	Prediction and Control with FA: Control with	Program. assignment (Function	

		approximation	approximation & control)	
13	Mon, Apr 8	Prediction and Control with FA: Policy Gradient	Quiz: Policy Gradient Method	Chapter 13, (pp. 321 - 336)
13	Wed, Apr 10	Prediction and Control with FA: Policy Gradient		
13	Fri, Apr 12	Prediction and Control with FA: Policy Gradient	Program. assignment (Average Reward Softmax Actor-Critic using Tile-coding)	
<b>Thu, Apr 25 at 9am</b>		<b>Final exam</b> (Students must verify this date on BearTracks when it is posted)		