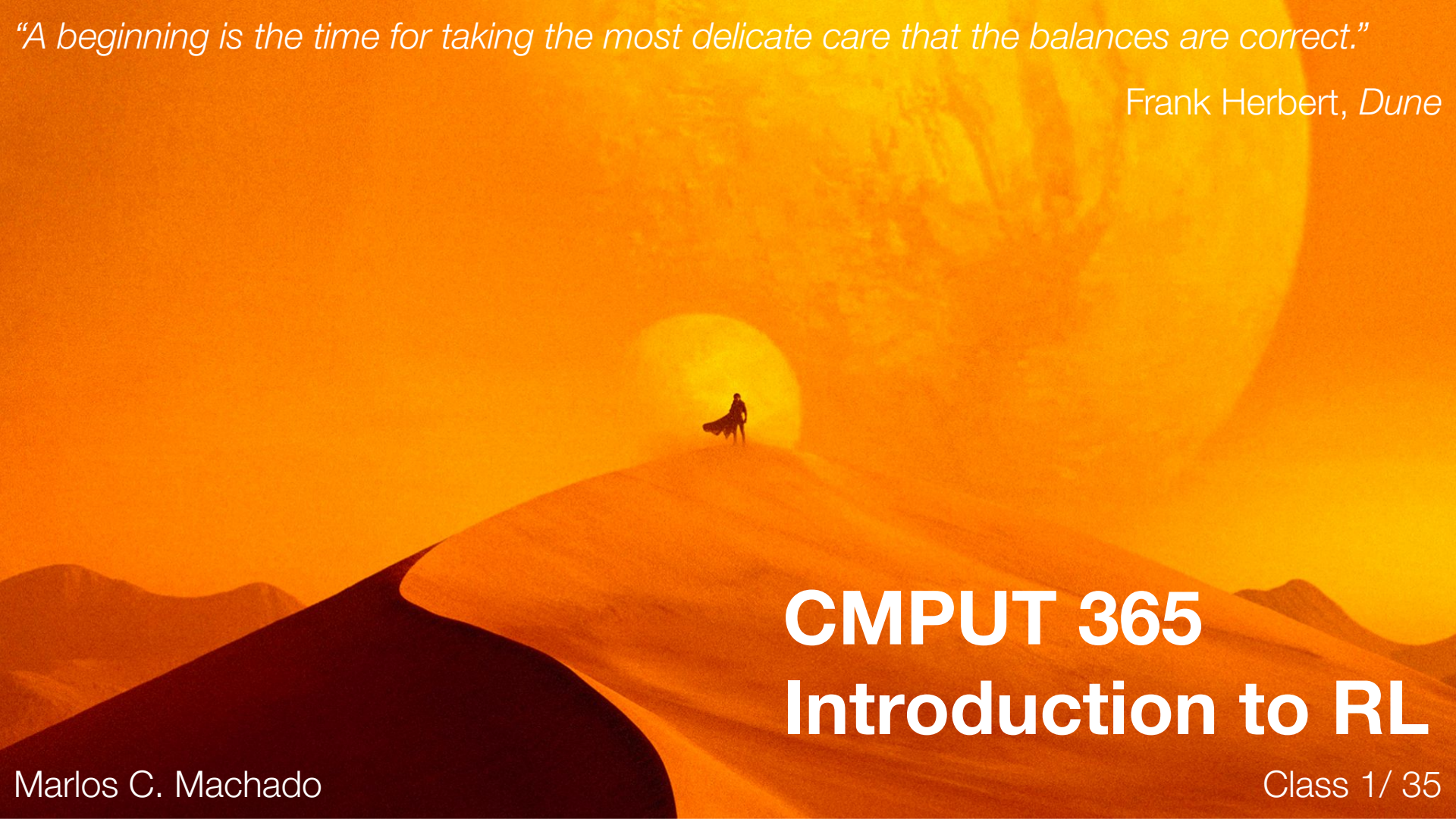
“A beginning is the time for taking the most delicate care that the balances are correct.”

Frank Herbert, *Dune*



CMPUT 365
Introduction to RL

Marlos C. Machado

Plan

- Introduction
- Course logistics
  - Instruction team
  - Pre-requisites
  - Textbook
  - Coursera
  - Academic integrity
  - Evaluation
- What is reinforcement learning?

Note

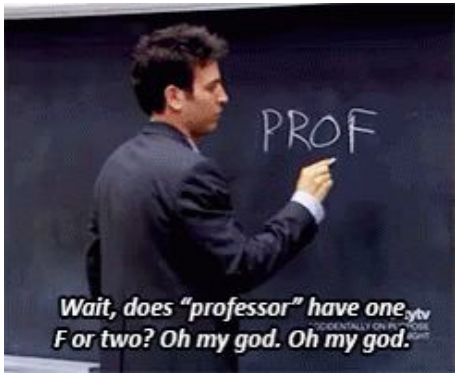
Lectures may be audio recorded for the purpose of a student's individual study as part of an approved academic accommodation.

Please, interrupt me at any time!

https://pngtree.com/freepng/question-expression-cartoon-illustration_4545209.ht

About myself

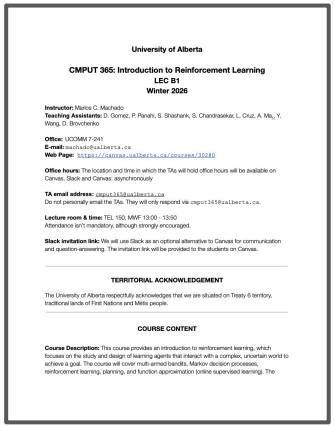
- Name: Marlos C. Machado
- I was born in Brazil
- I have been living in Edmonton for 10+ years
- I have 2 kids
- Ph.D. working on reinforcement learning
  - Interned at Microsoft Research, IBM Research, and DeepMind
- Worked 4 years at Google Brain and DeepMind
  - Among several other things, we deployed RL to fly balloons in the stratosphere



Course overview and logistics

Start here!

- Canvas: ***link***
- Slack: ***link***
- My website: ***link***
- Google drive: ***link***



University of Alberta

CMPUT 365: Introduction to Reinforcement Learning
LEC B1
Winter 2026

Instructor: Marlos C. Machado
Teaching Assistants: D. Gomez, P. Panahi, S. Shashank, S. Chandrasekar, L. Cruz, A. Mu, Y. Wang, D. Brovchenko

Office: UCOMM 7-241
E-mail: machado@ualberta.ca
Web Page: https://canvas.ualberta.ca/courses/30240

Office hours: The location and time in which the TAs will hold office hours will be available on Canvas. Slack and Canvas: asynchronously

TA email address: cmput365@ualberta.ca
Do not personally email the TAs. They will only respond via cmput365@ualberta.ca.

Lecture room & time: TEL 150, MWF 13:00 - 13:50
Attendance isn't mandatory, although strongly encouraged.

Slack invitation link: We will use Slack as an optional alternative to Canvas for communication and question-answering. The invitation link will be provided to the students on Canvas.

TERRITORIAL ACKNOWLEDGEMENT

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

COURSE CONTENT

Course Description: This course provides an introduction to reinforcement learning, which focuses on the study and design of learning agents that interact with a complex, uncertain world to achieve a goal. The course will cover multi-armed bandits, Markov decision processes, reinforcement learning, planning, and function approximation (online supervised learning). The

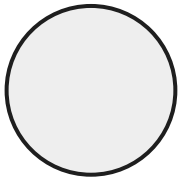
http://www.phdcomics.com/comics/archive/phd051013s.gif

Key resources

- Syllabus
  - Canvas, Slack, my website, Google Drive.
- Teaching assistants





Diego • Parham • Shashank • Siddarth • Lucas • Aaron • Yuyang • Dasha

- TA email address: cmput365@ualberta.ca
- My email address: machado@ualberta.ca
- Slack invitation link: ***link***

I want to make this course a **safe** and **inclusive** environment, for everyone.

It is ok to make mistakes.

We should all strive to be **respectful** to each other.

If you want me to address you by a **different name**, or if you want to tell me your **pronouns**, I'm more than happy to hear!

Office hours

- Slack and Canvas: Asynchronous
- Marlos: *After class* *@ here*
- TAs: *To be announced this week*

Syllabus [Canvas, Slack, website, Google Drive]

Pre-requisites

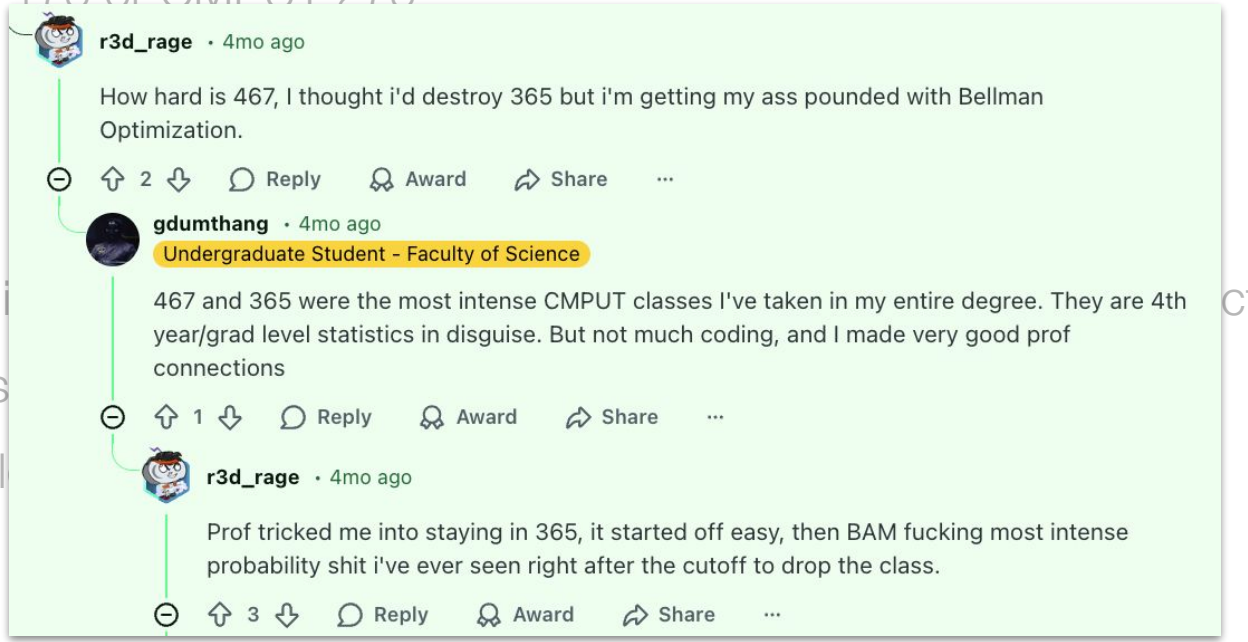
- CMPUT 175 or CMPUT 275
- CMPUT 267 or 466, or STAT 265

- Python
- Probability (e.g., expectations of random variables, conditional expectations)
- Calculus (e.g., partial derivatives)
- Linear algebra (e.g., vectors and matrices)

You should either be familiar with these topics or be ready to pick them up quickly as needed by consulting outside resources.

Pre-requisites

- CMPUT 175 or CMPUT 275
- CMPUT
- Python
- Probabili ctations)
- Calculus
- Linear al



to pick them up quickly as needed by consulting outside resources.

This will **not** be a flipped classroom!

- In the past, this course used to be taught in a flipped classroom
  - Roughly, you are initially introduced to **new topics outside** the classroom, using classroom time to explore topics in greater depth
- The number of students in this class has been steadily increasing, though
  - I don't know how to scale a flipped classroom without relying more and more on TAs to teach you
- Some of the feedback I received revolved around it feeling too repetitive
  - First read the textbook, watch the recorded lectures, do exercises, and then come to class
  - You can (and **should**) still do some of that before coming to class
- All this to say this will be a regular class, for better or for worse 😄

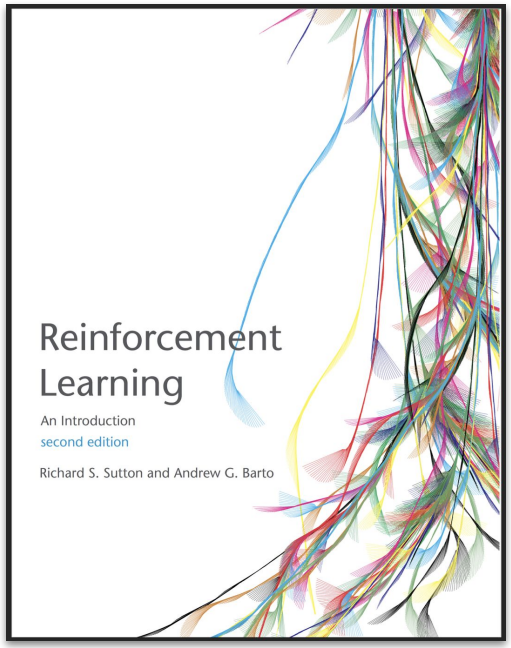
Required textbook

Reinforcement Learning: An Introduction

Richard S. Sutton & Andrew G. Barto

MIT Press. 2nd Edition.

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



- You will need to read the book!
  That's how you study for this course!
- The book is really good!

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	9 x 1% = 9%	Day of the last class on the topic of the week at 23:59:59 (see Course schedule at the end for details)
Assessments (graded quizzes/notebooks on Coursera)	9 x 2.5% = 22.5%	Day of the last class on the topic of the week at 23:59:59 (see Course schedule at the end for details)
Midterm 1 exam	20 %	January 30, 2026
Midterm 2 exam	20%	March 4, 2026
Final exam	30%	April 16, 2026*

Syllabus [Canvas, Slack, website, Google Drive]

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	9 x 1% = 9%	Day of the last class on the topic of the week at 23:59:59 (see Course schedule at the end for details)
Assessments (graded quizzes/notebooks on Coursera)	9 x 2.5% = 22.5%	Day of the last class on the topic of the week at 23:59:59 (see Course schedule at the end for details)
Midterm 1 exam	20 %	January 30, 2026
Midterm 2 exam	20%	March 4, 2026
Final exam	30%	April 16, 2026*

Syllabus [Canvas, Slack, website, Google Drive]

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	9 x 1% = 9%	Day of the last class on the topic of the week at 23:59:59 (see Course schedule at the end for details)
Assessments (graded quizzes/notebooks on Coursera)	9 x 2.5% = 22.5%	Day of the last class on the topic of the week at 23:59:59 (see Course schedule at the end for details)
Coursera, almost every week (starting next week, Jan 12): 31.5%		
Final exam	30%	April 16, 2026*

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	9 x 1% = 9%	Day of the last class on the topic of the week at 23:59:59 (see Course schedule at the end for details)
Assessments (graded quizzes/notebooks on Coursera)	9 x 2.5% = 22.5%	Day of the last class on the topic of the week at 23:59:59 (see Course schedule at the end for details)

Coursera, almost every week (starting next week, Jan 12): 31.5%

Late submissions will not be accepted. There are 11 quizzes and 11 graded assignments. You're expected to do all of them, but s**t happens, so you can miss 2 of each and still get full marks.

You need them all for the certificate!

Syllabus [Canvas, Slack, website, Google Drive]

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	9 x 1% = 9%	Day of the last class on the topic of the week at 23:59:59 (see Course schedule at the end for details)
Assessments (graded quizzes/notebooks on Coursera)	9 x 2.5% = 22.5%	Day of the last class on the topic of the week at 23:59:59 (see Course schedule at the end for details)
Midterm 1 exam	20 %	January 30, 2026
Midterm 2 exam	20%	March 4, 2026

Two midterms, summing to 40%. Closed book.

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	9 x 1% = 9%	Day of the last class on the topic of the week at 23:59:59 (see Course schedule at the end for details)
Assessments (graded quizzes/notebooks on Coursera)	9 x 2.5% = 22.5%	Day of the last class on the topic of the week at 23:59:59 (see Course schedule at the end for details)
Midterm 1 exam	20 %	January 30, 2026
Midterm 2 exam	20%	March 4, 2026

Two midterms, summing to 40%. Closed book.

If you miss the midterm, you can can apply for an excused absence. If granted, the weight of the missed midterm will be deferred to the final.

Syllabus [Canvas, Slack, Website, Google Drive]

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	9 x 1% = 9%	Day of the last class on the topic of the week at 23:59:59 (see Course schedule at the end for details)
Assessments (graded quizzes/notebooks on Coursera)	9 x 2.5% = 22.5%	Day of the last class on the topic of the week at 23:59:59 (see Course schedule at the end for details)
Midterm 1 exam	20 %	January 30, 2026
Midterm 2 exam	20%	March 4, 2026
Final exam	30%	April 16, 2026*

Syllabus [Canvas, Slack, website, Google Drive]

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	9 x 1% = 9%	Day of the last class on the topic of the week at 23:59:59 (see Course schedule at the end for details)
Assessments (graded quizzes/notebooks on Coursera)	9 x 2.5% = 22.5%	Day of the last class on the topic of the week at 23:59:59 (see Course schedule at the end for details)
Final exam	30%	April 16, 2026*

The final, worth 30%, will be about the whole course.
If you miss the final, you can apply to a deferred final examination.

Syllabus [Canvas, Slack, website, Google Drive]

GRADE EVALUATION

Assessment	Weight	Date
Practice quizzes (80% pass)	9 x 1% = 9%	Day of the last class on the topic of the week at 23:59:59 (see Course schedule at the end for details)
Assessments (graded quizzes/notebooks on Coursera)	9 x 2.5% = 22.5%	Day of the last class on the topic of the week at 23:59:59 (see Course schedule at the end for details)
Midterm 1 exam	20 %	January 30, 2026
Midterm 2 exam	20%	March 4, 2026
Final exam	30%	April 16, 2026*

Syllabus [Canvas, Slack, website, Google Drive]

GRADE EVALUATION

Assessment	Weight	D
Practice quizzes (80% pass)	9 x 1% = 9%	
Assessments (graded quizzes/notebooks on Coursera)	9 x 2.5% = 22.5%	
Midterm 1 exam	20 %	
Midterm 2 exam	20%	
Final exam	30%	

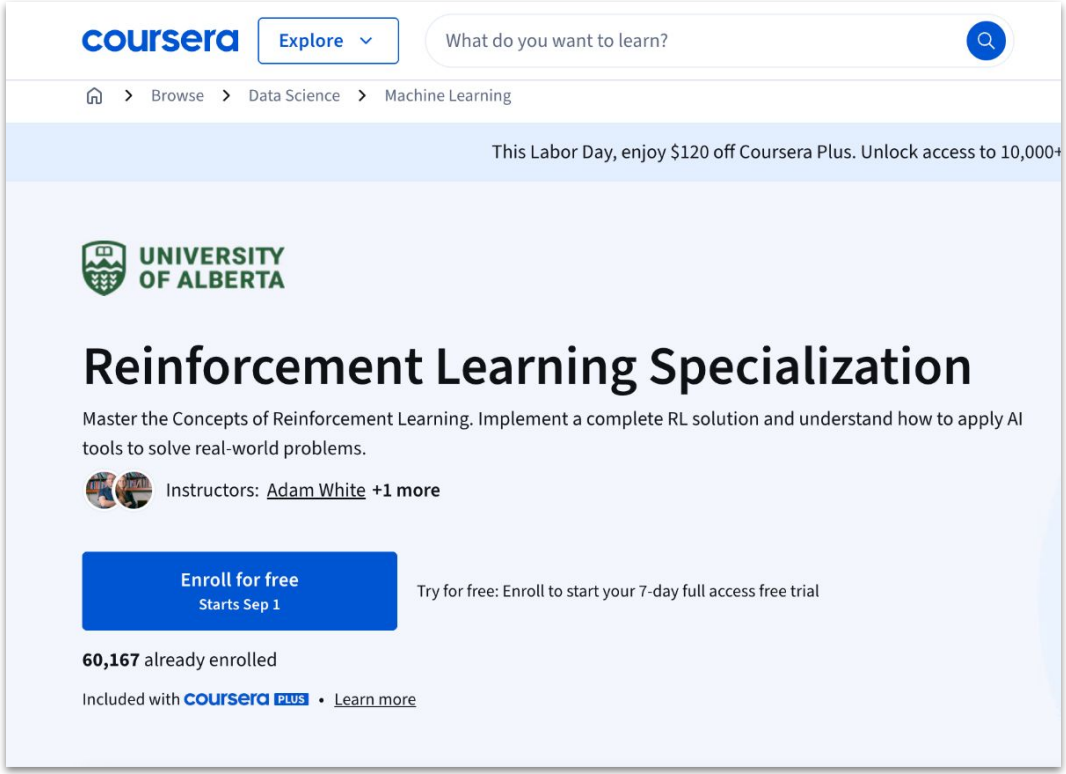
Total: **101.5%.**
You can **not-submit 2 quizzes** and **2 assessments.**

Grades **will not** be rounded at the end, and **no more extra marks** will be given. **No exceptions.**

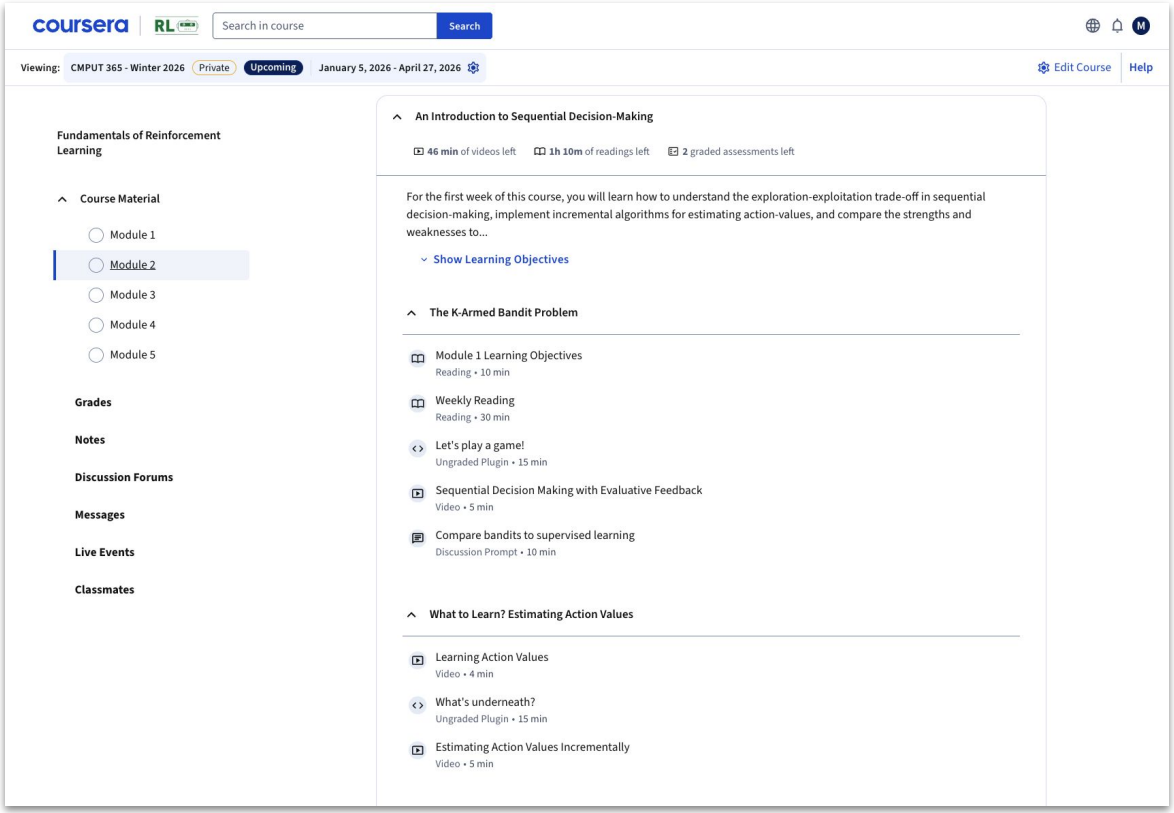
Syllabus [Canvas, Slack, website, Google Drive]

Coursera

- Coursera will be essential to CMPUT 365
- Today or tomorrow, you will be added to a private session of the RL courses (we will use your university's email)
  - If you don't have access you should let me know!
  - **IMPORTANT: If you don't use the private session you won't get credit for submitted work!**



Coursera


coursera
RL
Search in course
Search
Viewing:
CMPUT 365 - Winter 2026
Private
Upcoming
January 5, 2026 - April 27, 2026
Edit Course
Help
Fundamentals of Reinforcement Learning
Course Material
Module 1
Module 2
Module 3
Module 4
Module 5
Grades
Notes
Discussion Forums
Messages
Live Events
Classmates
An Introduction to Sequential Decision-Making
46 min of videos left
1h 10m of readings left
2 graded assessments left
For the first week of this course, you will learn how to understand the exploration-exploitation trade-off in sequential decision-making, implement incremental algorithms for estimating action-values, and compare the strengths and weaknesses to...
Show Learning Objectives
The K-Armed Bandit Problem
Module 1 Learning Objectives
Reading • 10 min
Weekly Reading
Reading • 30 min
Let's play a game!
Ungraded Plugin • 15 min
Sequential Decision Making with Evaluative Feedback
Video • 5 min
Compare bandits to supervised learning
Discussion Prompt • 10 min
What to Learn? Estimating Action Values
Learning Action Values
Video • 4 min
What's underneath?
Ungraded Plugin • 15 min
Estimating Action Values Incrementally
Video • 5 min

Academic integrity

- Code of Student Behaviour
- Student Conduct Policy
- Academic Integrity website

- **Appropriate collaboration:** You are 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.
- **Cell phones:** Cell phones are to be turned off during lectures, labs and seminars.
- **Recording and/or Distribution of Course Materials:** Audio or video recording, digital or otherwise, by students is allowed only with my prior written consent as a part of an approved accommodation plan.

Academic integrity – **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. Thus, in order to achieve such learning outcomes, you can submit each practice quiz and graded assignment multiple times, which allows for many learning opportunities.

The use of advanced AI-tools based on large-language models such as ChatGPT 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."*

Schedule

- The course will be structured in “weeks”. **Not every week starts on Monday**
- We have 12 weeks of content classes and we’ll cover 13 weeks of the MOOC
  - This corresponds to 9 chapters of the textbook

Schedule

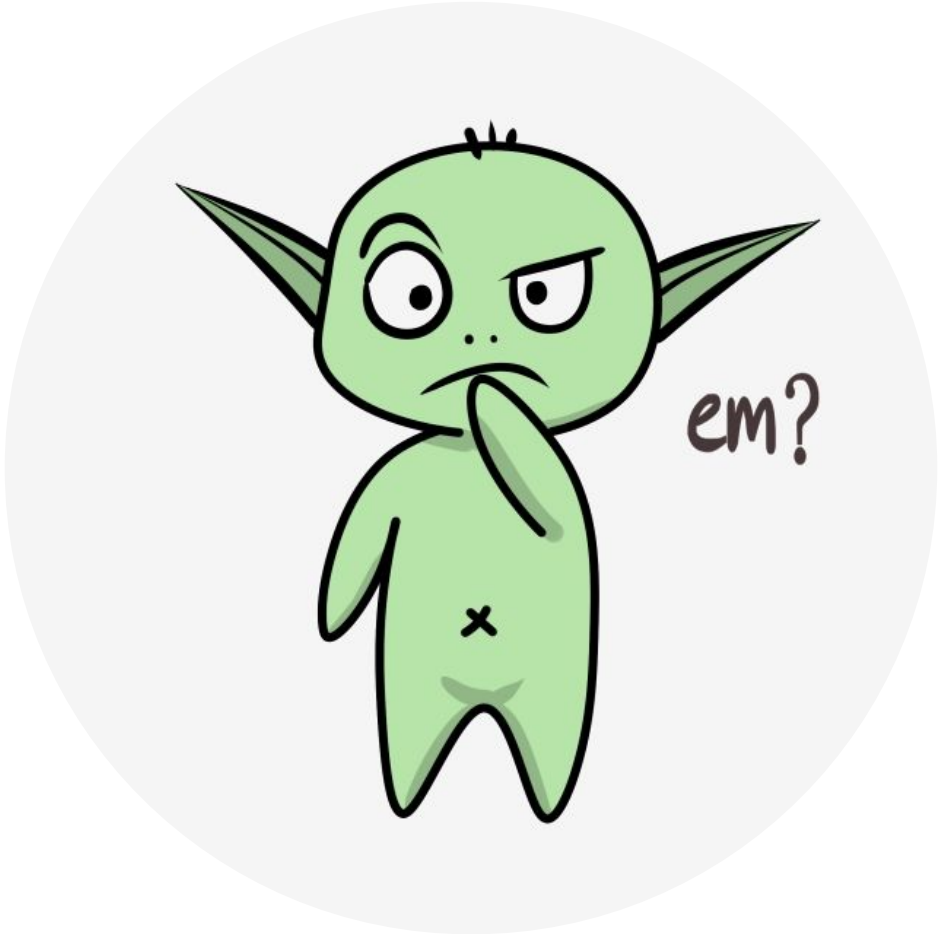
- The course will be structured in “weeks”. **Not every week starts on Monday**
- We have 12 weeks of content classes and we’ll cover 13 weeks of the MOOC
  - This corresponds to 9 chapters of the textbook
- A practice quiz and a graded assignment are due at the end of each “week” in terms of content – You should look at the syllabus / schedule all the time
- The deadline for submitting assignments and quizzes is 23:59:59

Schedule

Course Schedule & Assigned Readings

Week	Date	Topic	Deadlines (all due at 23:59:59)	Readings
0	Mon, Jan 5	Course overview Discussion about what is reinf. learning		
0	Wed, Jan 7	Background review: Probability, statistics, linear algebra, and calculus		
1	Fri, Jan 9	Fundamentals of RL: An introduction to sequential decision-making		Chapter 2, up to §2.7 (pp. 25-36), and §2.10 (pp. 42-44)
1	Mon, Jan 12	Fundamentals of RL: An introduction to sequential decision-making	Practice quiz and Progr. assignment (Bandits & exploration / exploitation)	
2	Wed, Jan 14	Fundamentals of RL: Markov decision processes (MDPs)		Chapter 3, up to §3.3 (pp. 47-56)
2	Fri, Jan 16	Fundamentals of RL: Markov decision processes (MDPs)	Practice quiz (MDPs)	
3	Mon, Jan 19	Fundamentals of RL: Value functions & Bellman equations		Chapter 3, §3.5-§3.8 (pp. 58-69)
3	Wed, Jan 21	Fundamentals of RL: Value functions & Bellman equations		
3	Fri, Jan 23	Fundamentals of RL: Value functions & Bellman equations	Practice and Graded quiz (Value functions & Bellman equations)	
4	Mon, Jan 26	Fundamentals of RL: Dynamic programming		Chapter 4, §4.1-§4.4 (pp. 73-84); §4.6-§4.7 (pp. 86-89)
4	Wed, Jan 28	Fundamentals of RL: Dynamic programming	Practice quiz and Progr. assignment (Optimal policies with dyn. progr.)	
	Fri, Jan 30	**Midterm exam 1**		
5	Mon, Feb 2	Sample-based learning methods: MC methods for Prediction & Control		Chapter 5, up to §5.5 (pp. 91-108); §5.10 (pp. 115-116)

Syllabus [Canvas, Slack, website, Google Drive]



https://pngtree.com/freepng/question-expression-cartoon-illustration_4545209.ht

What is reinforcement learning?

Artificial intelligence

Artificial intelligence

"AI is the ability of machines to perform tasks that are typically associated with human intelligence, such as learning and problem-solving." –Wikipedia

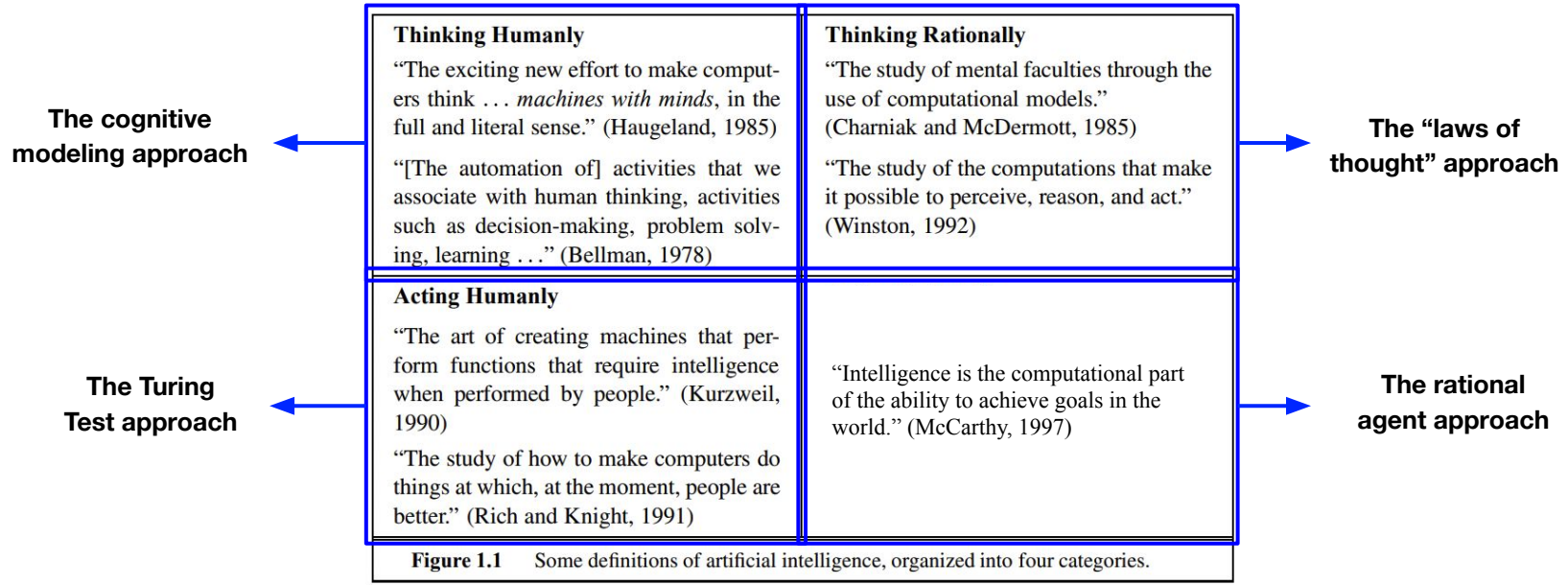


Figure 1.1 Some definitions of artificial intelligence, organized into four categories.

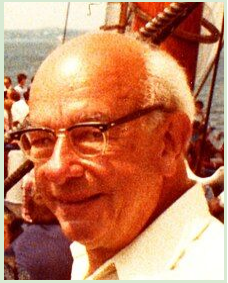
(Russell & Norvig; 2010)

Artificial intelligence

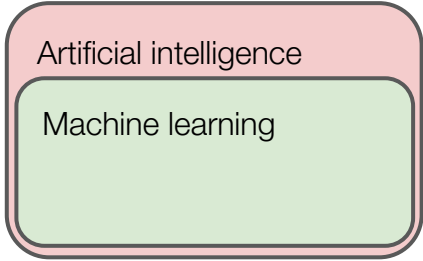
Artificial intelligence

“AI is the ability of machines to perform tasks that are typically associated with human intelligence, such as learning and problem-solving.” –Wikipedia

> The less a science has advanced, the more its terminology tends to rest on an uncritical assumption of mutual understanding.
>
> – W. V. Quine

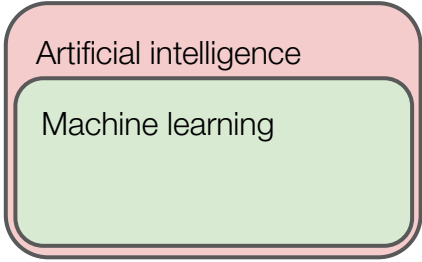


https://upload.wikimedia.org/wikipedia/commons/c/cd/Willard_Van_Orman_Quine_on_Bluenose_II_in_Halifax_NS_harbor_1980.



Machine learning

Machine learning is a subfield of AI in which the system's desired behavior is not explicitly programmed, instead it is *learned* from data



Machine learning

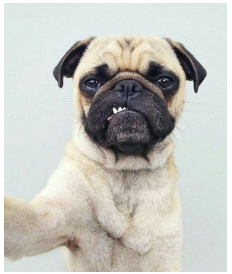
Machine learning is a subfield of AI in which the system's desired behavior is not explicitly programmed, instead it is *learned* from data

- "*Supervised learning* is learning from a training set of labeled examples provided by a knowledgeable external supervisor" (Sutton & Barto; 2018)

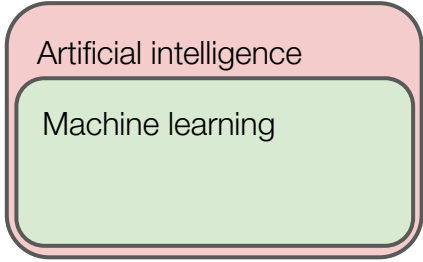
Cat

Cat

Not cat



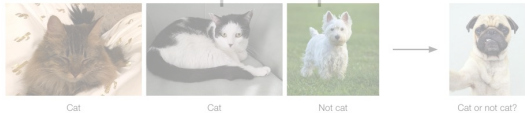
Cat or not cat?



Machine learning

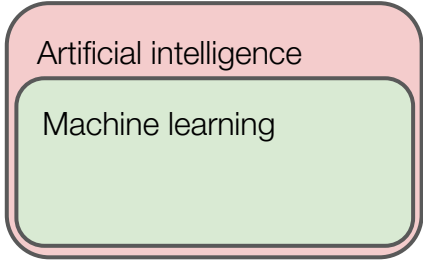
Machine learning is a subfield of AI in which the system's desired behavior is not explicitly programmed, instead it is *learned* from data

- "*Supervised learning* is learning from a training set of labeled examples provided by a knowledgeable external supervisor" (Sutton & Barto; 2018)



- "*Unsupervised learning* is typically about finding structure hidden in collections of unlabeled data" (Sutton & Barto; 2018)

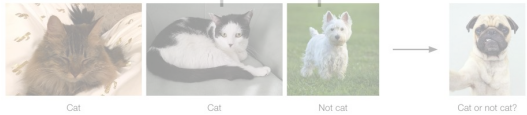
https://www.researchgate.net/publication/309515576/figure/fig1/AS:960348551122945@1605976345743/uits-can-be-clustered-in-different-ways-images-are-from-Internet-a-Fruits-b.gif



Machine learning

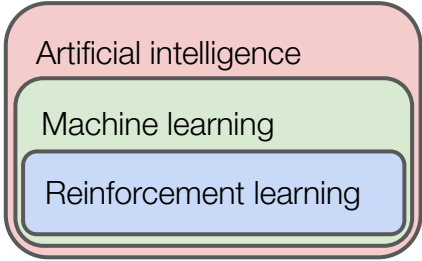
Machine learning is a subfield of AI in which the system's desired behavior is not explicitly programmed, instead it is *learned* from data

- "*Supervised learning* is learning from a training set of labeled examples provided by a knowledgeable external supervisor" (Sutton & Barto; 2018)



- "*Unsupervised learning* is typically about finding structure hidden in collections of unlabeled data" (Sutton & Barto; 2018)

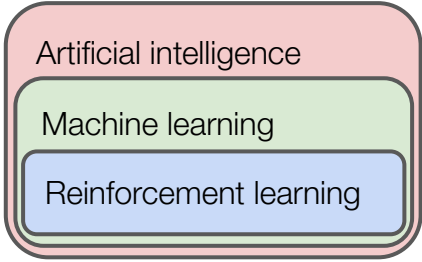
… and *reinforcement learning*!



Reinforcement learning

Reinforcement learning is a computational approach to learning from interaction to maximize a numerical reward signal (Sutton & Barto; 2018)

https://images.squarespace-cdn.com/content/v1/5bdf4a4d7c93279afce735f7/1548721628539-YNH1GBN9MXX3CQMCSYVR/treat+dog+in+background.jpg?format

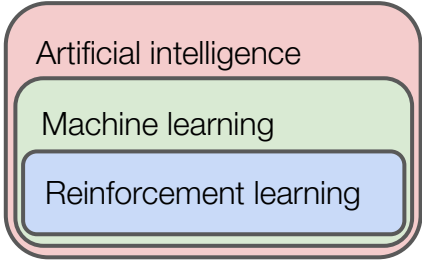


Reinforcement learning

Reinforcement learning is a computational approach to learning from interaction to maximize a numerical reward signal (Sutton & Barto; 2018)

- The idea of learning by interacting with our environment is very natural
- It is based on the idea of a learning system that wants something, and that adapts its behavior to get that

https://images.squarespace-cdn.com/content/v1/5bdf4a4d7c93279afce735f7/1548721628539-YNH1GBN9MXX3CQMCSYVR/treat+dog+in+background.jpg?format



Reinforcement learning

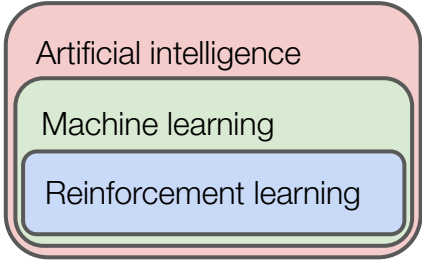
Reinforcement learning is a computational approach to learning from interaction to maximize a numerical reward signal (Sutton & Barto; 2018)

- The idea of learning by interacting with our environment is very natural
- It is based on the idea of a learning system that wants something, and that adapts its behavior to get that

Some features are unique to reinforcement learning:

- Trial-and-error
- The trade-off between exploration and exploitation
- The delayed credit assignment / delayed reward problem

https://images.squarespace-cdn.com/content/v1/5bdf4a4d7c93279afce735f7/1548721628539-YNH1GBN9MXX3CQMCSYVR/treat+dog+in+background.jpg?format



Reinforcement learning

Reinforcement learning is a computational a g from interaction to maximize a numerical reward signal (Sutto

- The idea of learning by interactin ent is very natural
- It is based on the idea of a at wants something, and that ac get that

Some features are un nt learning:

- Trial-and-erro
- The trade- ation and exploitation
- The delayed ment / delayed reward problem

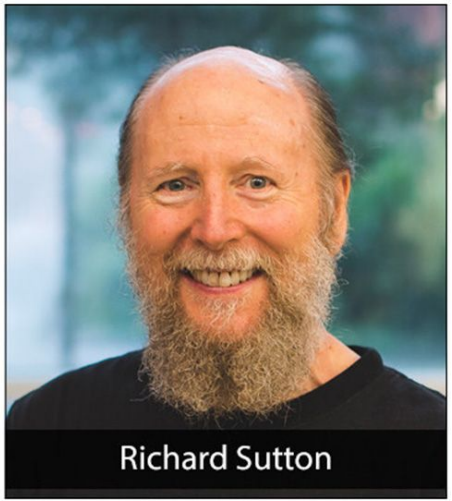
Problem or solution?

https://images.squarespace-cdn.com/content/v1/5bdf4a4d7c93279afce735f7/1548721628539-YNH1GBN9MXX3CQMCSYVR/treat+dog+in+background.jpg?format

RL is now commonly deployed in the real-world

- **Recommendation systems**
  - Ads, news articles, videos, etc
- **General game playing**
  - Go, Chess, Shogi, Atari 2600, Starcraft, Minecraft, Gran Turismo
- **Industrial automation**
  - Cooling commercial buildings
  - Inventory management
  - Gas turbine optimization
  - Optimizing combustion in coal-fired power plants
- **Algorithms**
  - Video compression on YouTube
  - Faster matrix multiplication
  - Faster sorting algorithms
- **Control / Robotics**
  - Navigating stratospheric balloons
  - Plast control for nuclear fusion
- **And more (see Csaba's slides)**
  - COVID-19 border testing
  - Conversational agents
  - ...

The 2024 ACM A.M. Turing Award Winners “Created” RL



Association for Computing Machinery

On intelligence, AGI, ASI, etc etc…

- People in the field have different, non-competing, perspectives and motivations
  - Some study RL to learn about / develop tools for solving sequential decision-making problems
  - Some look at RL as a computational model of intelligence

On intelligence, AGI, etc etc…

- People in the field have different, non-competing, perspectives and motivations
  - Some study RL to learn about / develop tools for solving sequential decision-making problems
  - Some look at RL as a computational model of intelligence
- RL was originally developed to understand intelligence/the brain
  - We should develop a critical view around these topics, and an ability to recognize hype / PR pieces

A. Barto (2024)

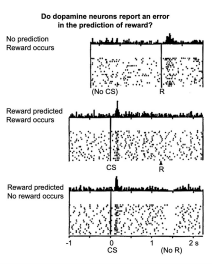
On intelligence, AGI, etc etc…

- People in the field have different, non-competing, perspectives and motivations
  - Some study RL to learn about / develop tools for solving sequential decision-making problems
  - Some look at RL as a computational model of intelligence

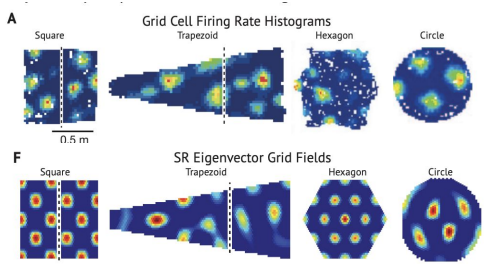
A. Barto (2024)

- RL was originally developed to understand intelligence/the brain
  - We should develop a critical view around these topics, and an ability to recognize hype / PR pieces
- Both perspectives are valid and both had had successes in the past

But they are different!!



(Schultz, Dayan, & Montague; 1997)

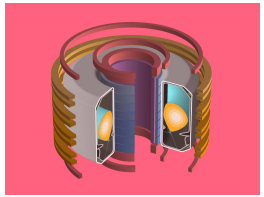


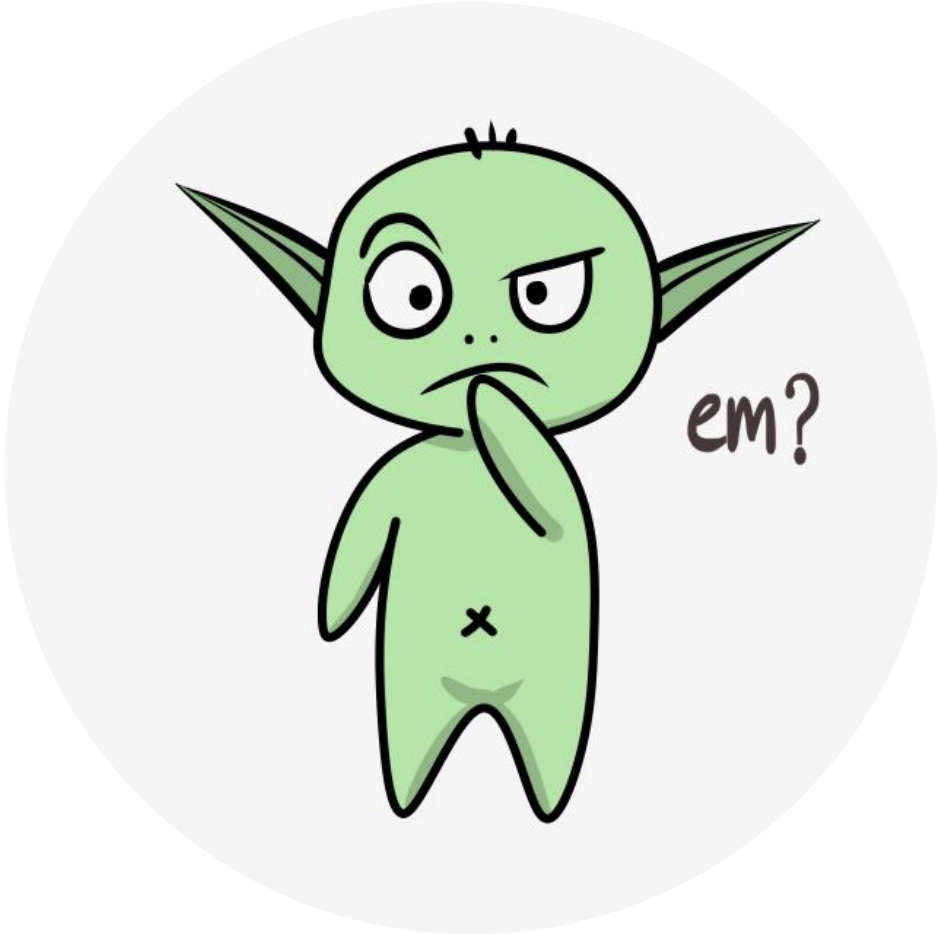
(Stachenfeld, Botvinick, & Gershman; 2017)

(Silver et al.; 2016)

(Bellemare et al.; 2020)

(Degrave et al.; 2022)





https://pngtree.com/freepng/question-expression-cartoon-illustration_4545209.ht

Next class

- What **I** plan to do:
  - Fundamentals of RL: An introduction to sequential decision-making (Bandits)

- What I recommend **YOU** to do for next class:
  - Make sure you have access to Coursera, Canvas, and Slack
  - Read Chapter 1 of the textbook (not mandatory)