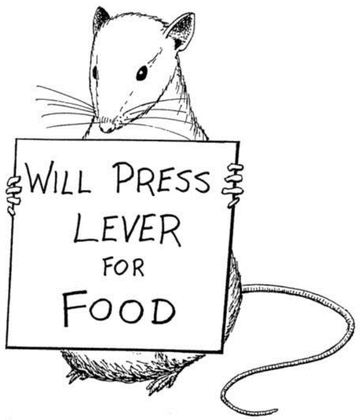
UCI University of California, Irvine

CS 277: Control and Reinforcement Learning

Winter 2026

Lecture 1: Introduction

Roy Fox
Department of Computer Science
School of Information and Computer Sciences
University of California, Irvine



Today's lecture

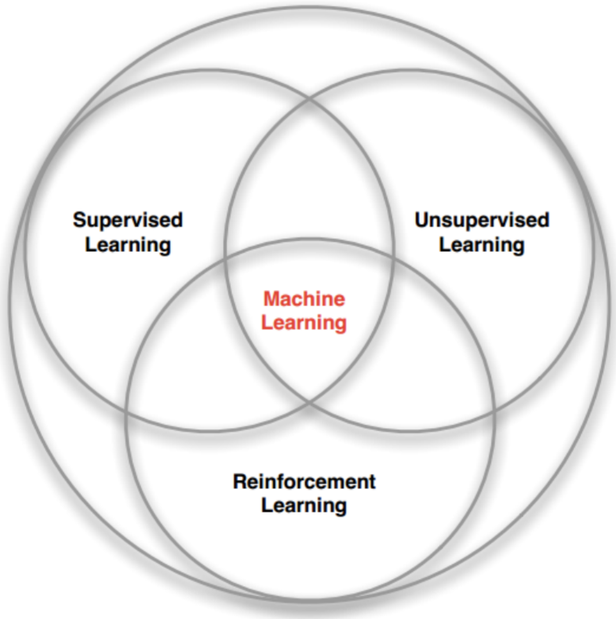
What is reinforcement learning?

Course logistics

Why is RL interesting?

RL ⊆ control learning ⊆ ML

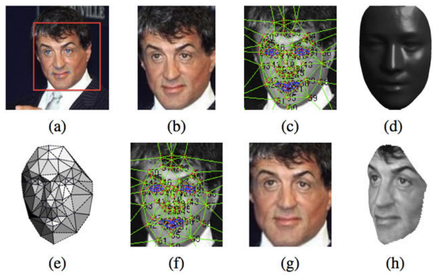
- Reinforcement Learning = learning from reinforcement (rewards)
  - But it came to encompass many settings of learning to control
  - Distinguished by data-driven sequential decision making
- Many consider RL a separate ML paradigm, but it can involve:
  - Supervised learning
  - Unsupervised learning
  - Active learning
  - Online learning



What is machine learning

- Can we build “intelligent” machines? Intelligence = good decision making
- Learning = taking in information to “know” more than you did before
- Machine learning = use data to make better decisions than before [Mitchell 1997]
- ML can help when other AI methods fail:
  - Experts are scarce
  - Rules / logic are hard to specify
  - Search space is too large
  - Models are unknown / hard to specify

Face recognition



Speech synthesis

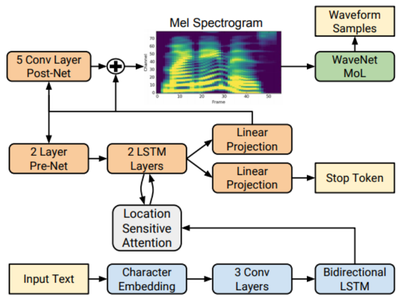
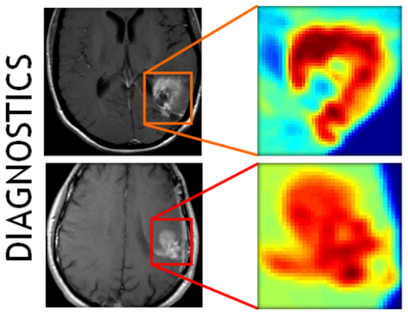


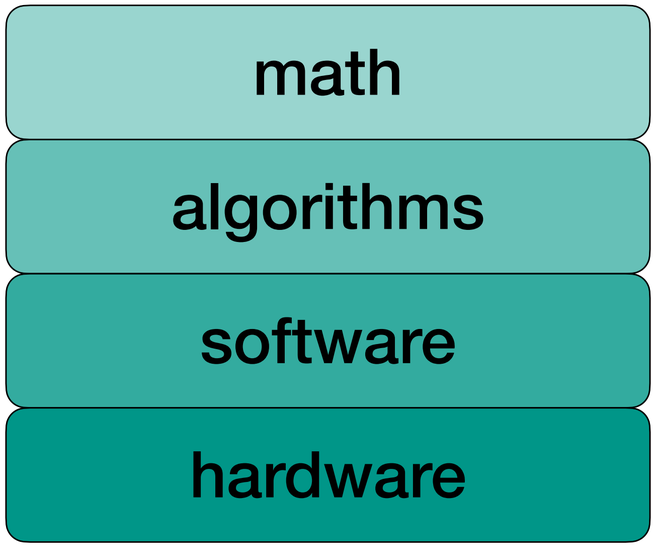
Fig. 1. Block diagram of the Tacotron 2 system architecture.

Medical diagnosis



[Taigman et al., 2014; Shen et al., 2018]

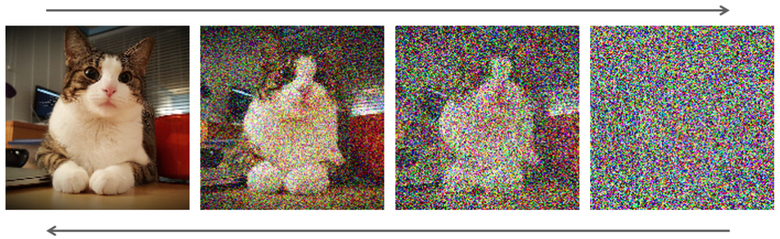
The ML stack



- Math: probability theory, (linear) algebra, computational learning theory
- Algorithms: ML algorithms, optimization, data structures
- Software: ML frameworks, databases, evaluation, deployment
- Hardware: cloud computing, distributed systems, cyber-physical systems

ML success stories

Image generation



Language generation

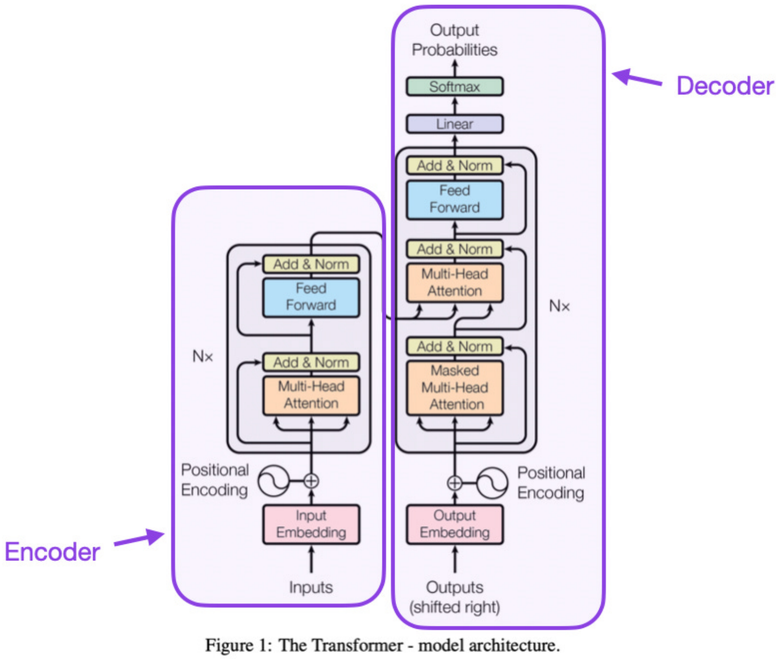
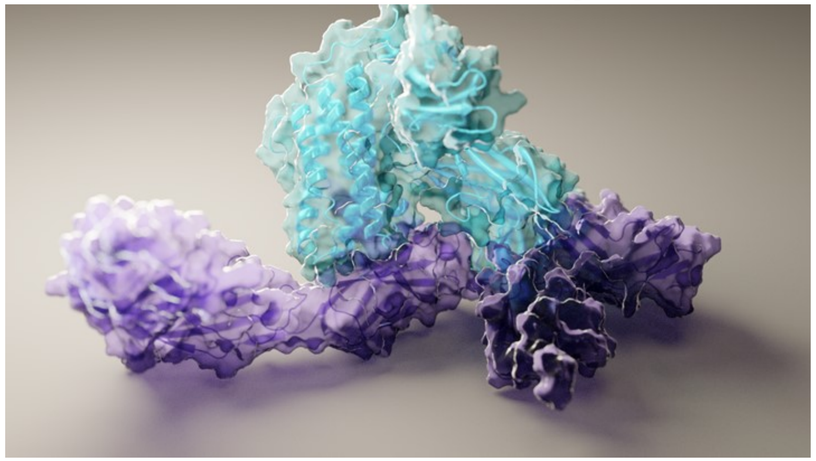


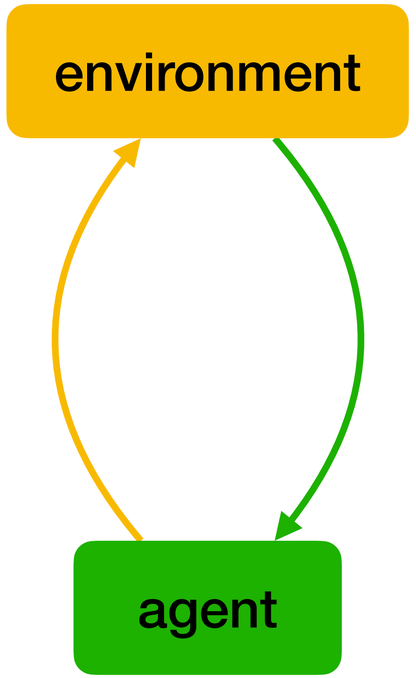
Figure 1: The Transformer - model architecture.

Protein folding



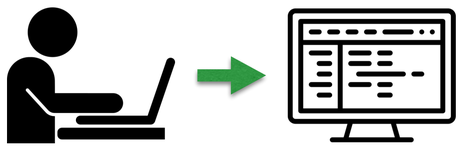
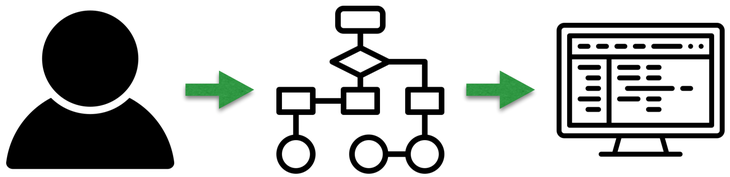
What is control learning (CL)?

- Intelligence appears in interaction with a complex system, not in isolation
  - An agent interacting with an environment
- Control = sequential decision making
  - Sense environment state s
  - Take action a
  - Repeat



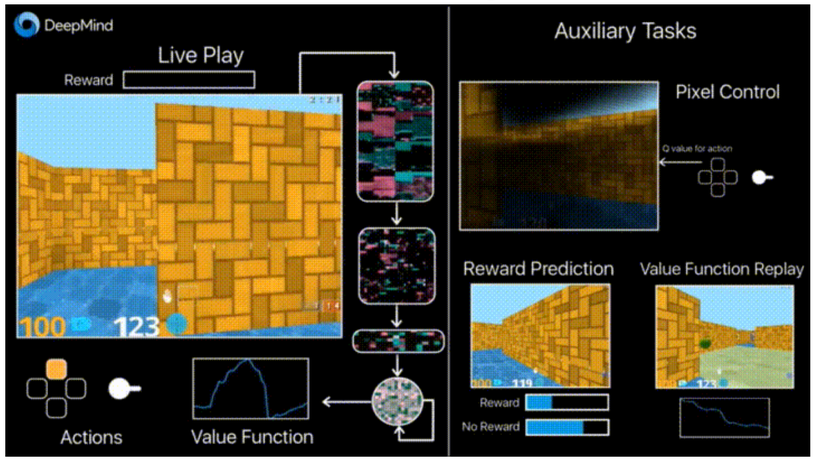
- Success can be measured by matching good actions — imitation learning (IL)
  - Or by accumulating high rewards $r(s, a)$ — reinforcement learning (RL)

Control preference elicitation

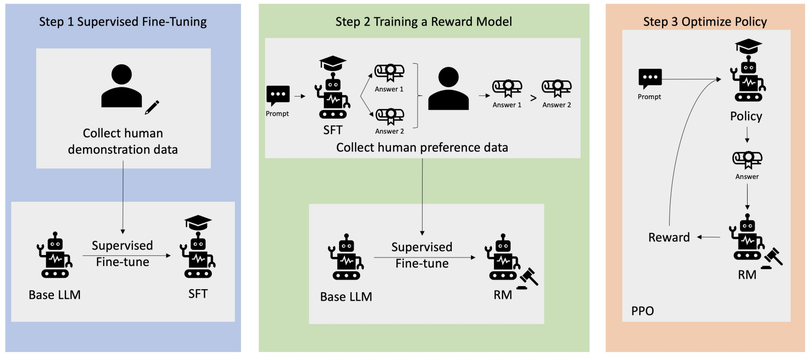
	Explicit	Implicit
"how"	**Programming** 	**Imitation Learning** 
"what"	**Instruction Following** 	**Reinforcement Learning** 

RL success stories

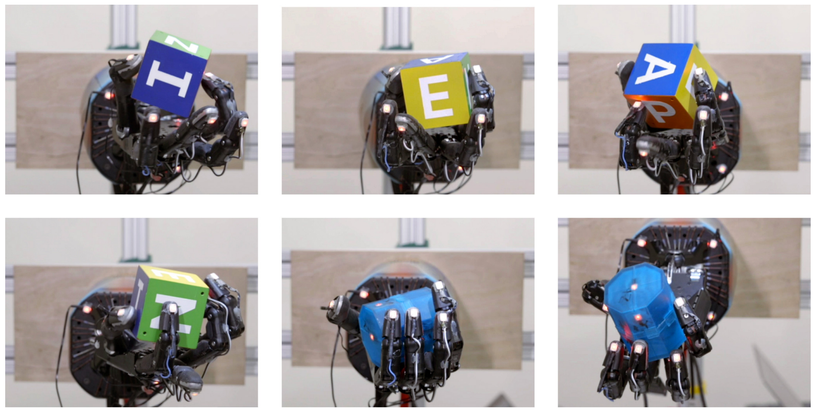
Spatial navigation



Generator fine-tuning



Dextrous manipulation



RL is ML... but special

- In RL, unlike supervised, no ground truth, only feedback (online learning)
- Exploration = the learner collects data by interaction
  - The agent decides on which states to train (active learning) — and test!
  - Cannot avoid some train–test mismatch
- Sequential decision making need to be coordinated
  - Optimization space is teeming with local optima
- A good policy may require memory
  - Agent state is latent → combine control and inference

Today's lecture

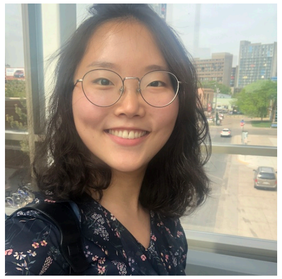
What is reinforcement learning?

Course logistics

Why is RL interesting?

Course logistics: general

- Course website: https://royf.org/crs/CS277/W26
  - Schedule; recordings; exercises; resources
- Forum: https://edstem.org/us/courses/90858
  - Announcements; discussions
- Office hours: in-person or on zoom
  - Welcome to schedule 15-min slots; individually or with classmates
- TA: Kyungmin Kim
  - Office hours: https://calendar.app.google/QQsruJxq9PF1CGcT6



Course logistics: lectures and discussions

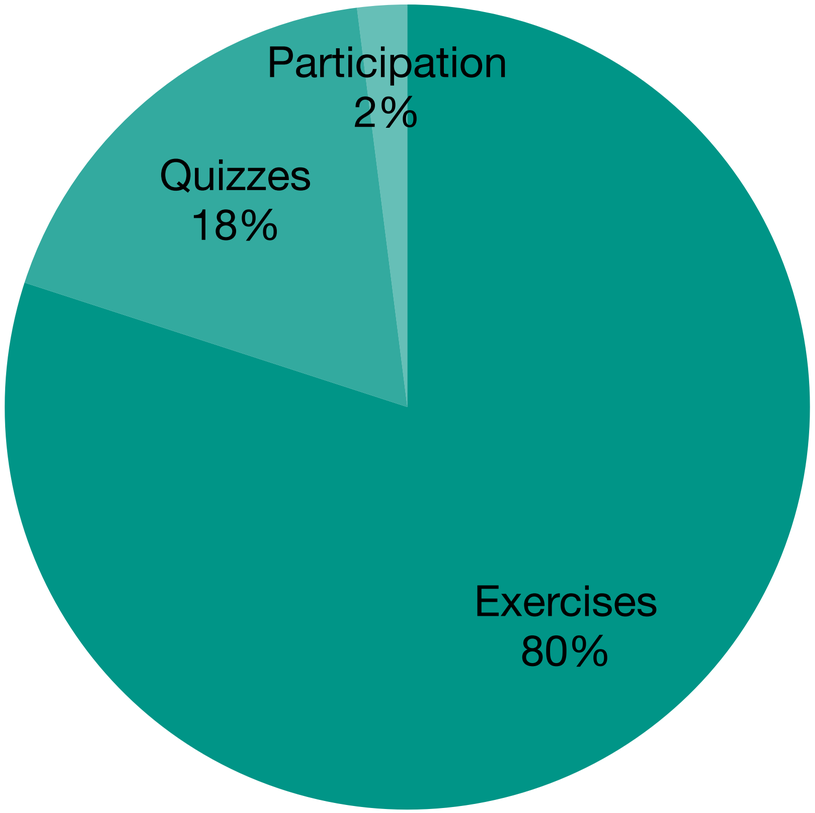
- Lectures
  - When: Tuesdays and Thursdays, 5–6:20pm
  - Where: ICS 180
  - Recorded when possible, uploaded to the course website
  - Attendance is optional but recommended
- Class discussions
  - Reviewing quizzes and exercises following deadline
  - Recaps, deep dives, freeform discussions

Course logistics: quizzes and exercises

- Quizzes
  - Weekly, about that week's topics; deadlines the following Monday
  - Discussed the following Tuesday in class
- Exercises
  - Roughly every other week; deadlines typically Friday
  - Understand RL concepts; apply RL techniques in Python
  - Discussed the following Thursday in class
- Submission: https://www.gradescope.com/courses/1210041

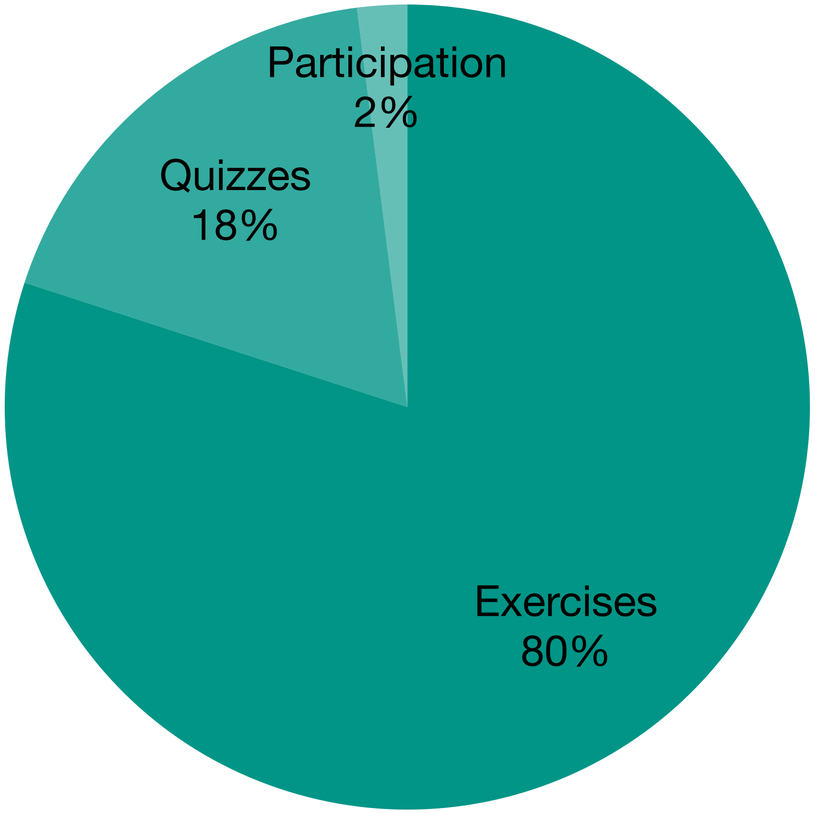
Grading policy: exercises

- Show your math, code, and results
- Encouraged to discuss with me or classmates
  - But solve yourself
- 4 best of 5 exercises count for 20% each
- 5% bonus for scoring at least 50% on all 5
- Late submission: 5 grace days total



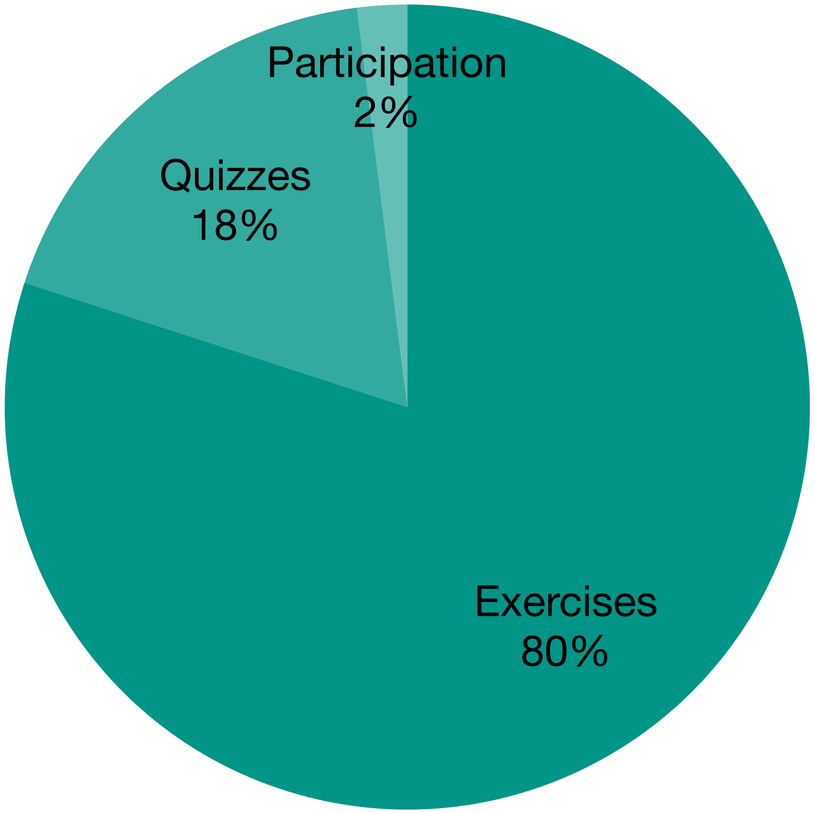
Grading policy: quizzes

- Review the week's topics, think about them a bit
- Aiming for 9 quizzes at 2% each = 18%
  - Half the score for submitting a complete quiz
  - Half the score for doing better than random guess
- No late submission



Grading policy: participation

- Class, office hours, or forum participation: 2%
  - Ask questions if you have any
  - Answer quiz or forum questions if you can
  - Share thoughtful comments
  - Post relevant useful links
  - Be on-topic (excluding administrative)
- Course evaluations: 2% bonus



What will it take to do well?

- We'll rely heavily on math: probability theory, linear algebra, calculus
  - I'm here to help, but solid background expected

- You'll need to code well in Python
- Some ideas are challenging — ask early what you don't fully understand
  - There'll be a lot going on, and nobody understands everything immediately
  - If you walk away with a good general understanding of the basics — that’s a win!
- Help your friends and get help — from me too — but never cheat!

Today's lecture

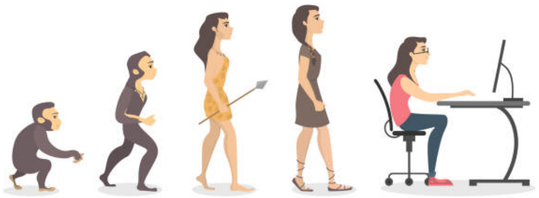
What is reinforcement learning?

Course logistics

Why is RL interesting?

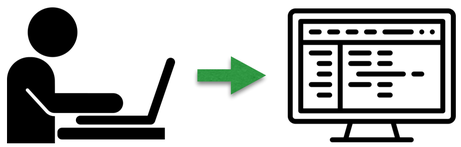
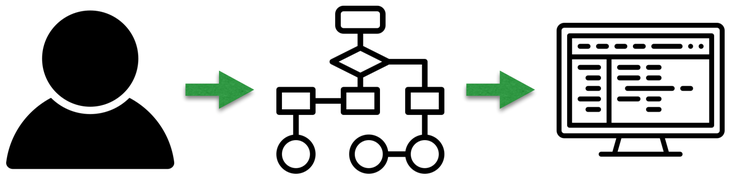
Why is RL powerful?

- Many (all?) problems can be formulated as control
  - But consider: is it sequential? multi-agent? a more specific structure?
- Active + online = very little supervision
  - Even incidental, like in evolution! Supervisor can be “surprised”



- More general CL: incorporate stronger supervision
  - Supervisor burden is a tradeoff between data amount ↔ informativeness

Control preference elicitation

	Explicit	Implicit
"how"	**Programming** 	**Imitation Learning** 
"what"	**Instruction Following** 	**Reinforcement Learning** 

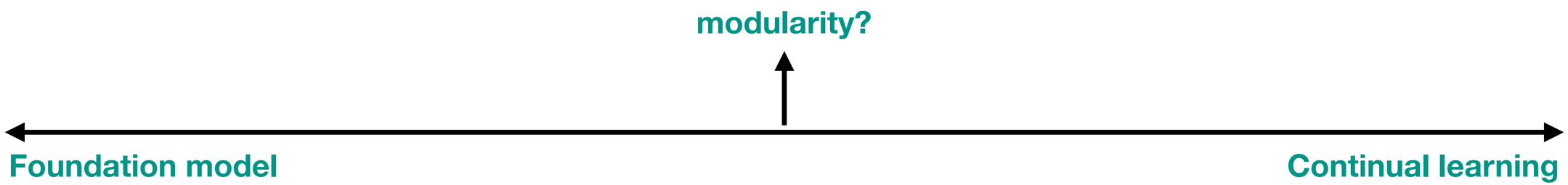
How is RL different?

online learning

	Provided correct output	Provided feedback
Teacher selects input	**IL from demonstrations** 	**Offline RL** 
Learner selects input	**IL from corrections** 	**RL** 

active learning

What would “solving” RL look like?



- Foundation model?
  - Large model
  - Huge amount of data
  - Centrally trained
  - Fine-tuned, built into pipelines
- Continual learning?
  - Flexible model
  - Ad-hoc (“on-task”) data
  - Distributed learning
  - Mixed supervision, shared learning

The last ML frontier?

Why is RL hard?

- It's all about the data: amount and informativeness

	Provided correct output	Provided feedback
Teacher selects input	**IL from demonstrations**  expert, train–test mismatch	**Offline RL**  extreme train–test mismatch
Learner selects input	**IL from corrections**  hard to give exploration	**RL**  weak signal, exploration

Logistics

logistics

- Follow announcements and discussions on ed
- See website for schedule, recordings, resources, etc.

assignments

- Quiz 1 due next Monday
- Exercise 1 to be published soon, due next Friday