This lecture will be recorded!

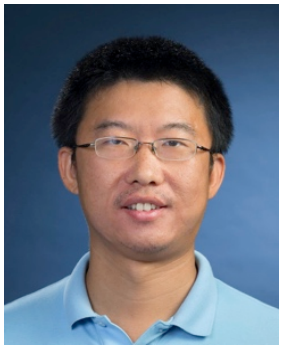
Welcome to

DS595: Reinforcement Learning --Introduction & Logistics

Prof. Yanhua Li

Time: 6:00pm –8:50pm Wednesday
Zoom Lecture
Spring 2022

Who am I?



Yanhua Li, PhD
Associate Professor,
Computer Science & Data Science

Joined WPI since 2015 Fall.
PhD, Computer Science, U of Minnesota, 2013
PhD, Electrical Engineering, BUPT, 2009

Research Interests: Big data analytics, Artificial Intelligence, Spatio-temporal Data Mining, Smart Cities;

Industrial Experience: Bell-Labs, Microsoft Research

http://users.wpi.edu/~yli15/index.html

Teaching Assistant

Yingxue Zhang
PhD Student with WPI Data Science Program

What is *this course* about?

- A advanced DS/CS/RBE course (primarily) for graduates
  - CS/DS/RBE Ph.D students in AI, DM, ML and related areas;
  - then, other Ph.D students or MS students with
    - Experience in Machine Learning, or equivalent knowledge.
    - Sufficient programming experience in python is expected so that you are comfortable to undertake the course projects.
    - You will have access to Ace, a WPI teaching cluster

    https://arc.wpi.edu/cluster-documentation/build/html/index.html

Topics for today

- What is reinforcement learning?
- Difference from Supervised and unsupervised machine learning?
- Application stories.
- Topics to be covered in this course.
- Course logistics

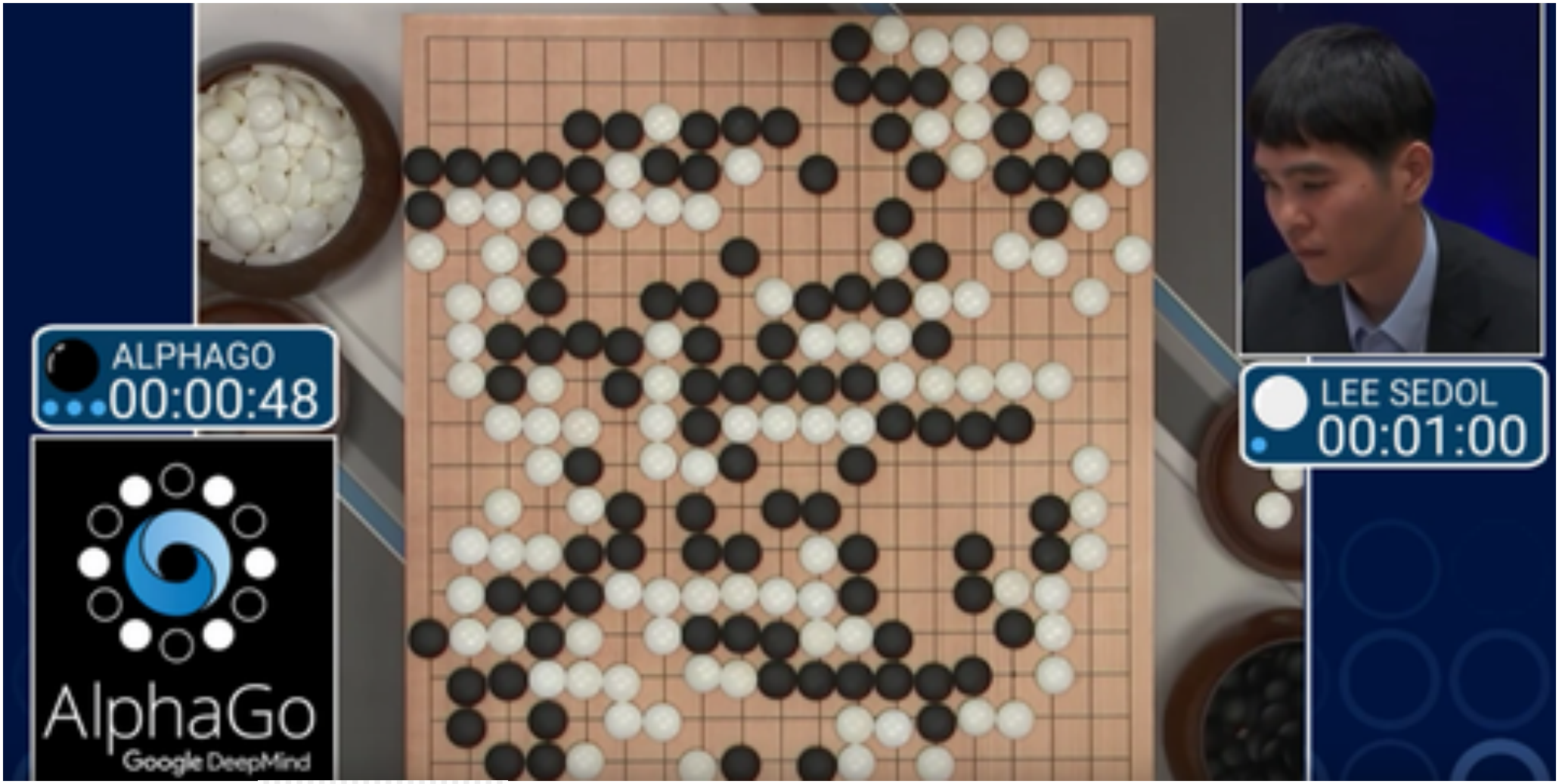
?

Reinforcement Learning
What can it do?

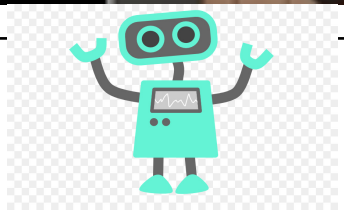
?????

Let’s see some more examples

Why (Deep) Reinforcement Learning?



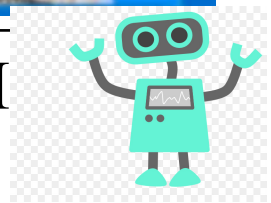
AlphaGo
Mar. 2016



Why (Deep) Reinforcement Learning?

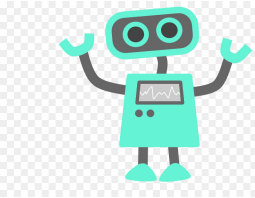
AlphaStar: Mastering the Real-Time Strategy Game StarCraft II

Apr. 2019



Why (Deep) Reinforcement Learning?

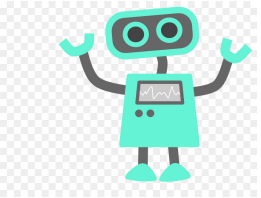
MineRL competition: Minecraft ObtainDiamond task.
Jun-Oct. 2019. http://minerl.io/competition/



Beyond Games -> Intelligent Agents

Autonomous vehicles

Intelligent and autonomous agents are as good as or doing better than human.

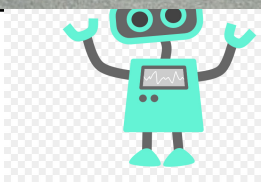


Beyond Games -> Robot Control

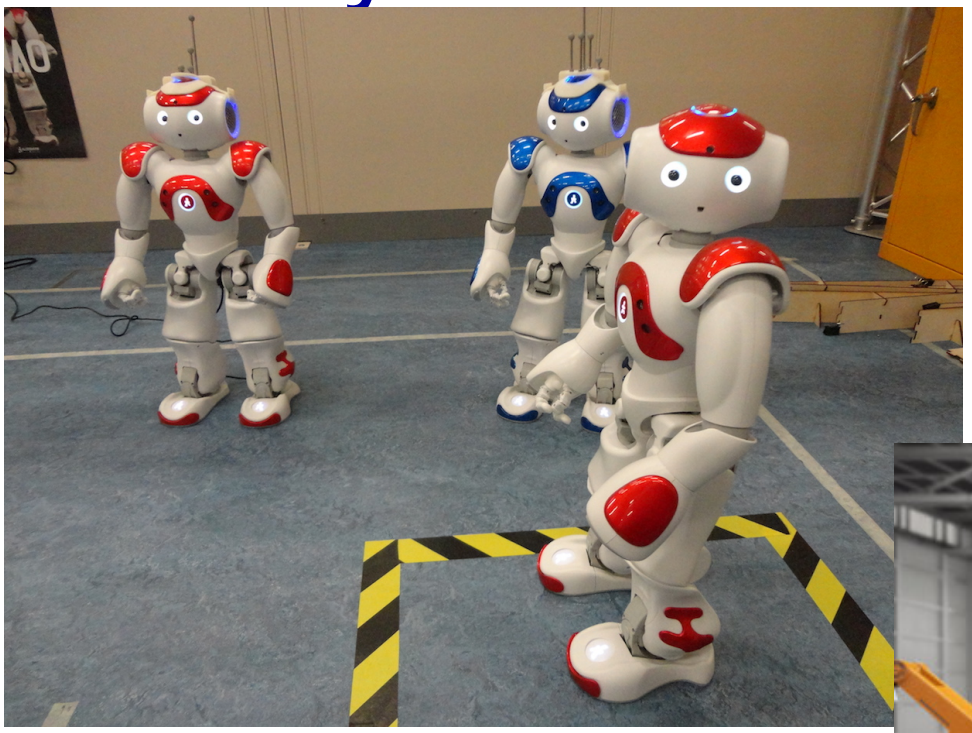
Drone Control

Unmanned Aircraft

Intelligent and autonomous agents are as good as or doing better than human.



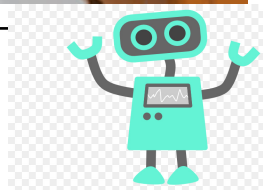
Beyond Games -> Robot Control



Robot Control

Industrial Robots

Intelligent and autonomous agents are as good as or doing better than human.



?

Reinforcement Learning
What is it?

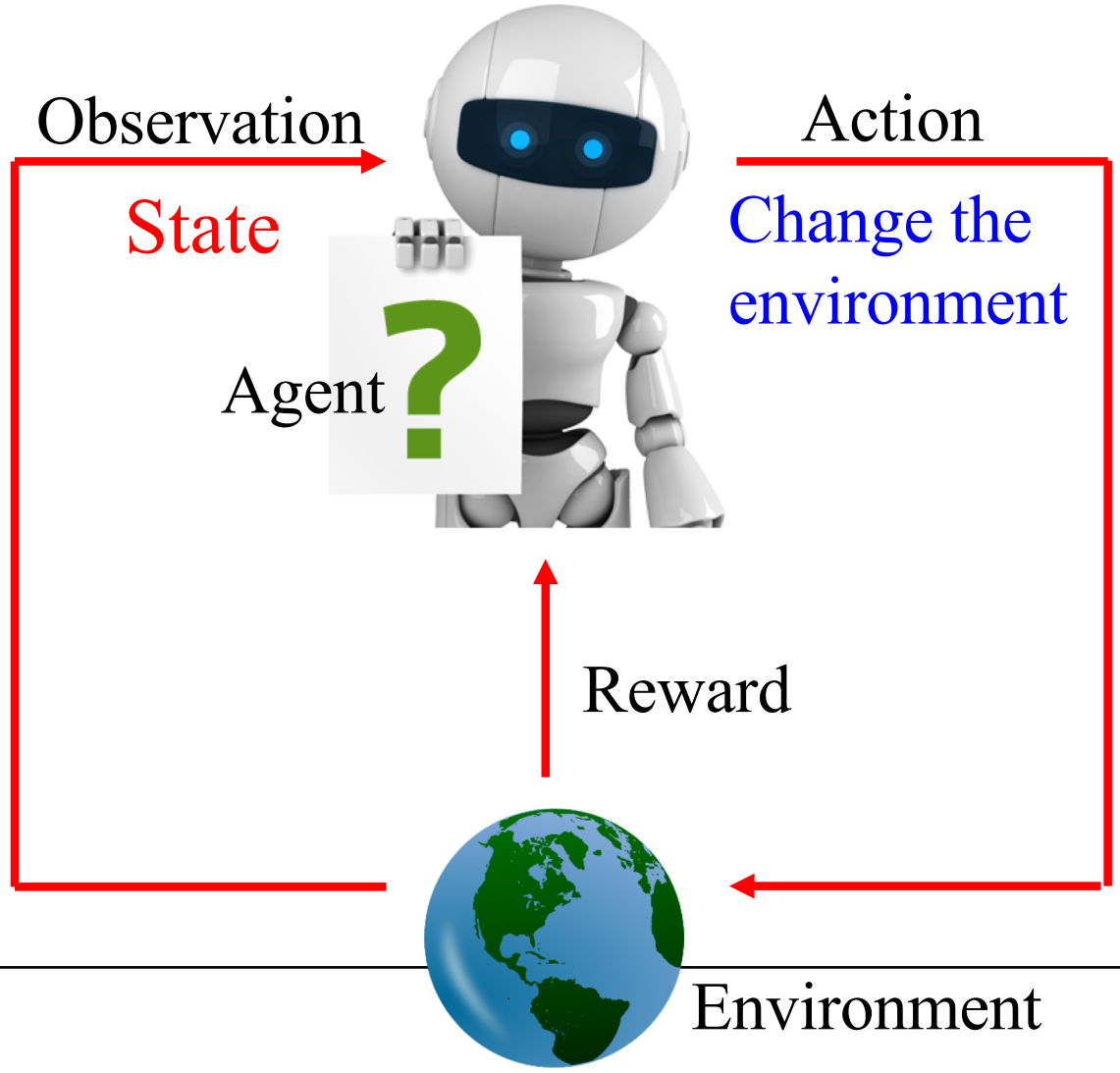
Training intelligent agents?

Reinforcement Learning
What is it?

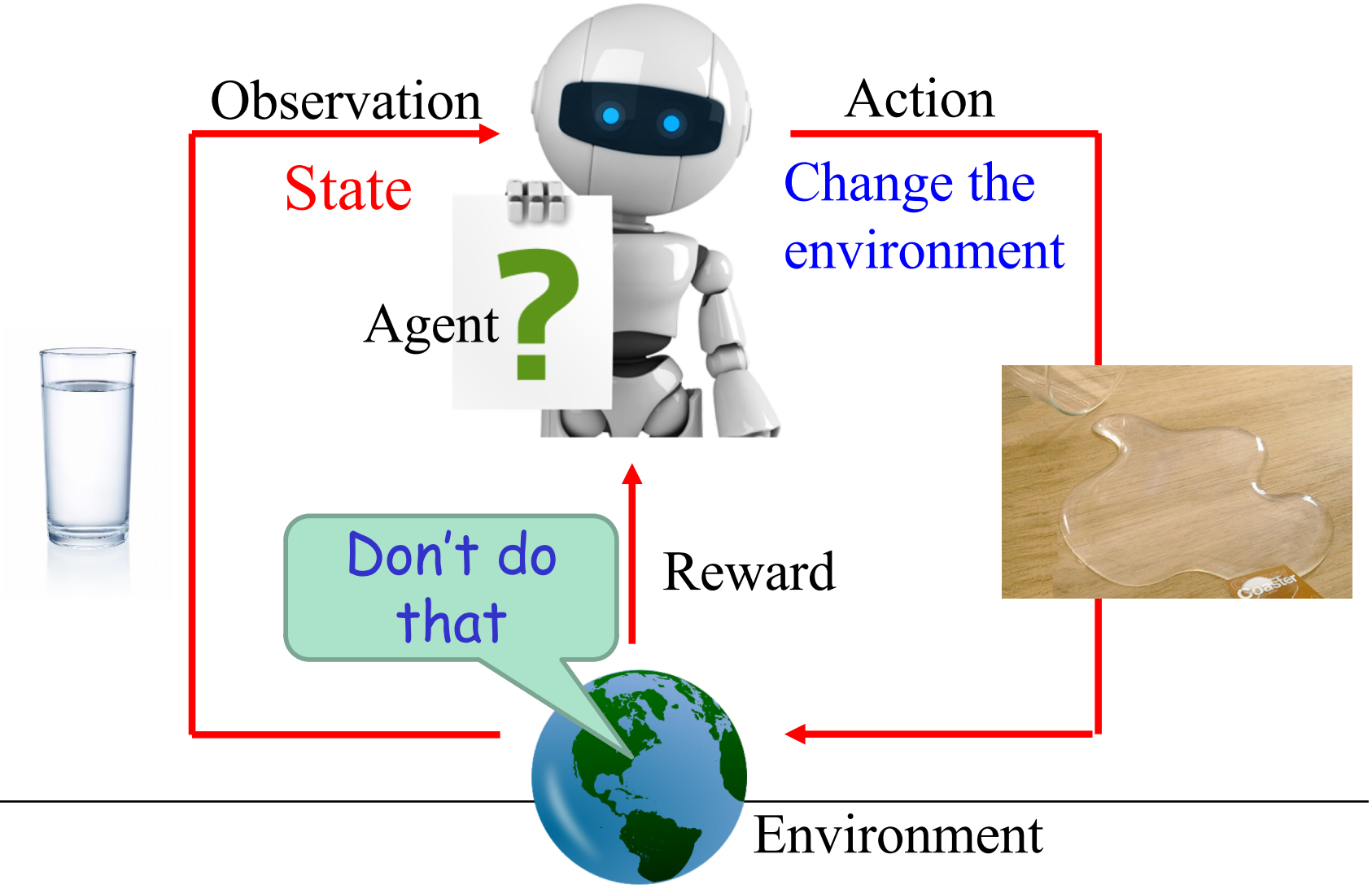
Reinforcement learning (RL) is an area of machine learning concerned with how software agents ought to take actions in an environment to maximize some notion of cumulative reward.

(From Wikipedia)

Scenario of Reinforcement Learning



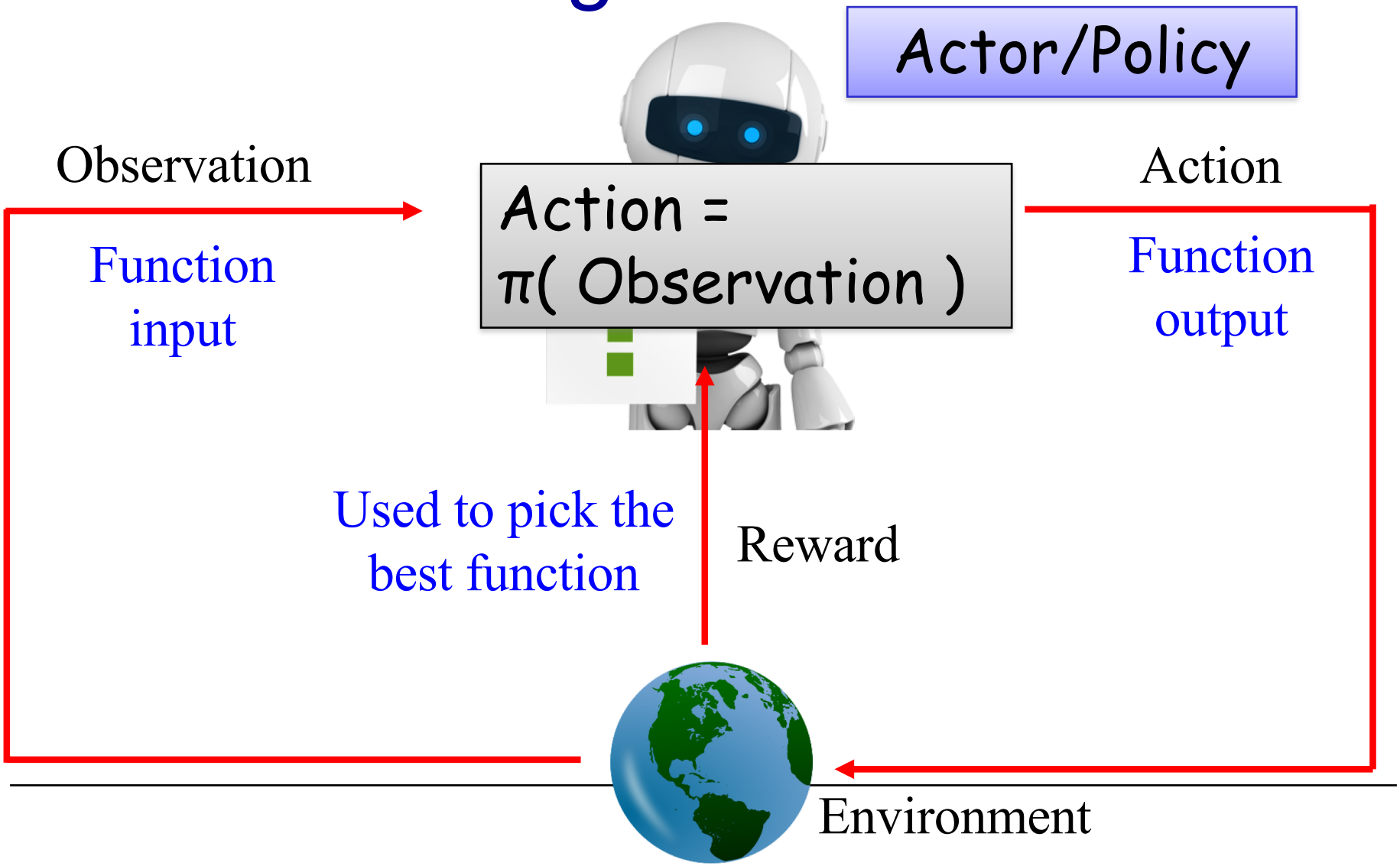
Scenario of Reinforcement Learning



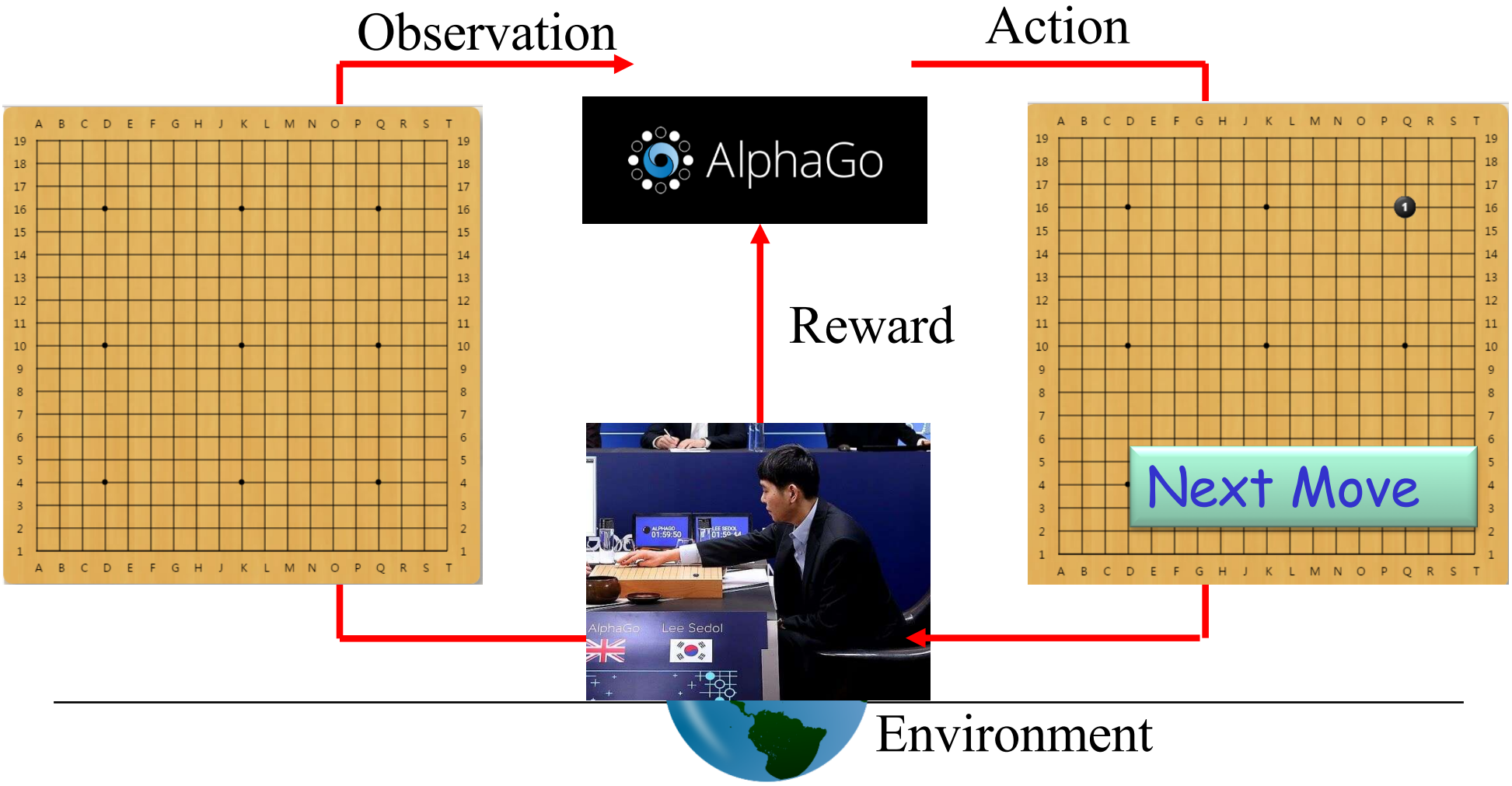
Scenario of Reinforcement Learning

https://yoast.com/how-to-clean-site-structure/

Reinforcement Learning ≈ Looking for a Function

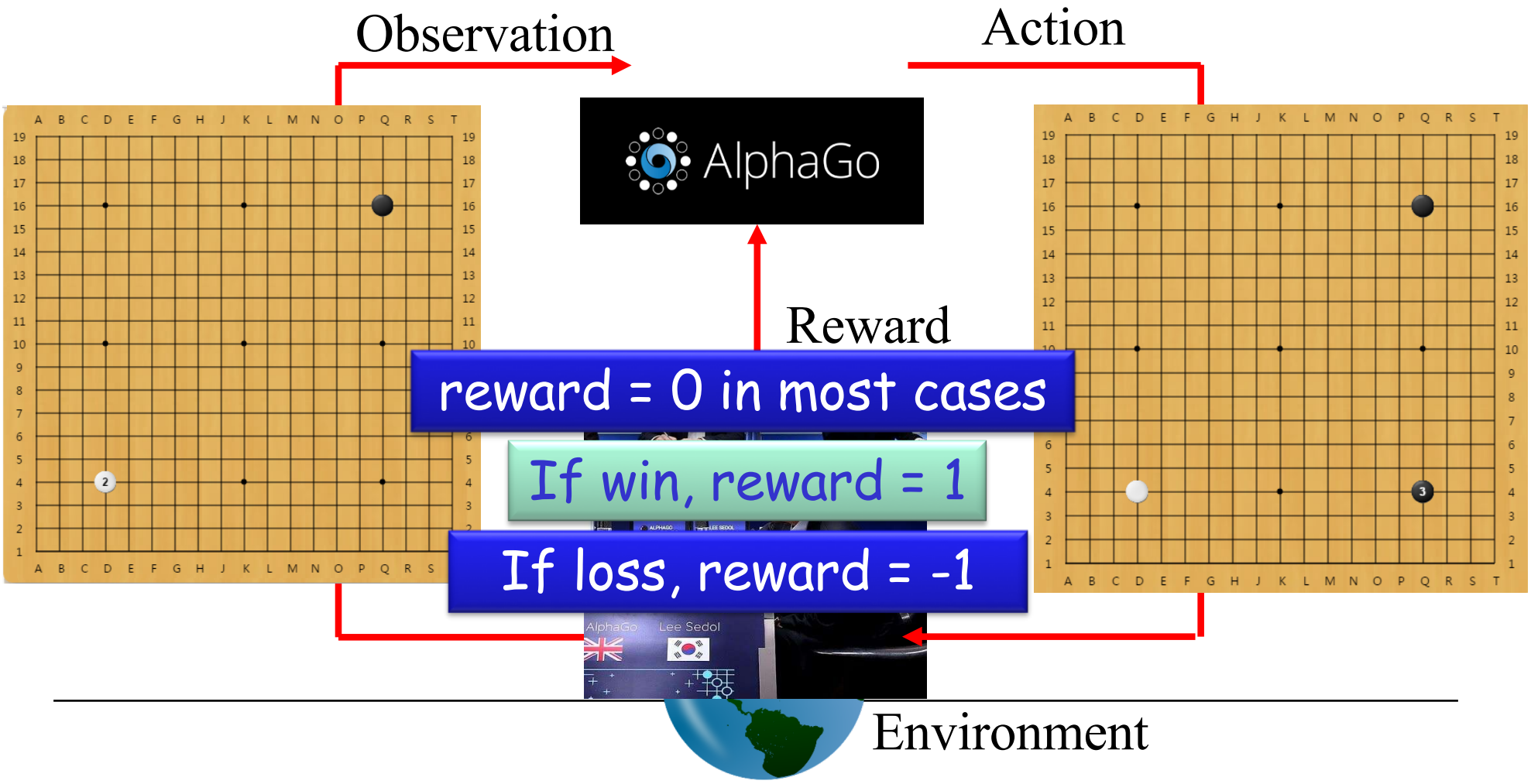


Learning to play Go



Learning to play

Agent learns to take actions maximizing expected reward.



Example: Playing Video Games

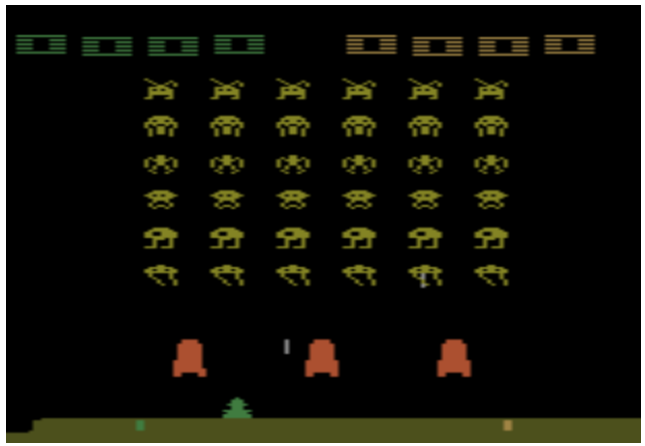
- Space invader

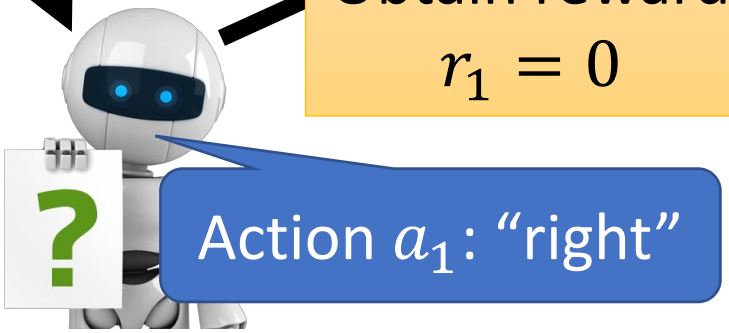
Example: Playing Video Game

Start with observation s_1

Observation s_2

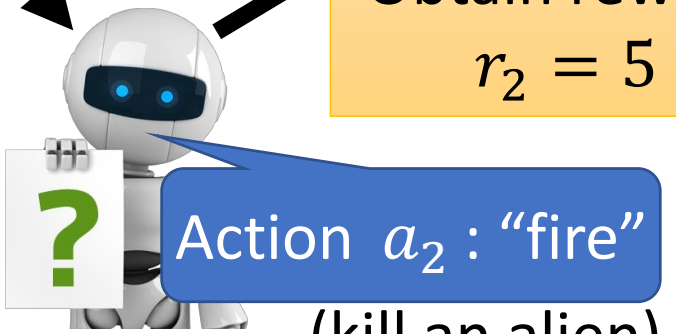
Observation s_3





Obtain reward $r_1 = 0$

Action a_1: "right"



Obtain reward $r_2 = 5$

Action a_2: "fire"

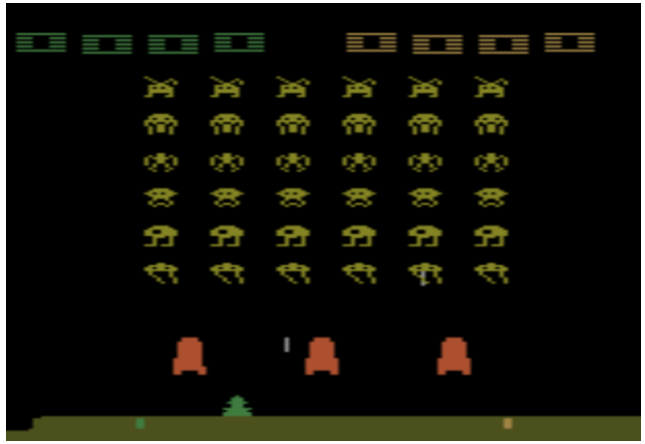
(kill an alien)

Usually there is some randomness in the environment

Example: Playing Video Game

Start with observation s_1

Observation s_2



Observation s_3

After many turns

Obtain reward r_T

Action a_T

This is an ***episode***.

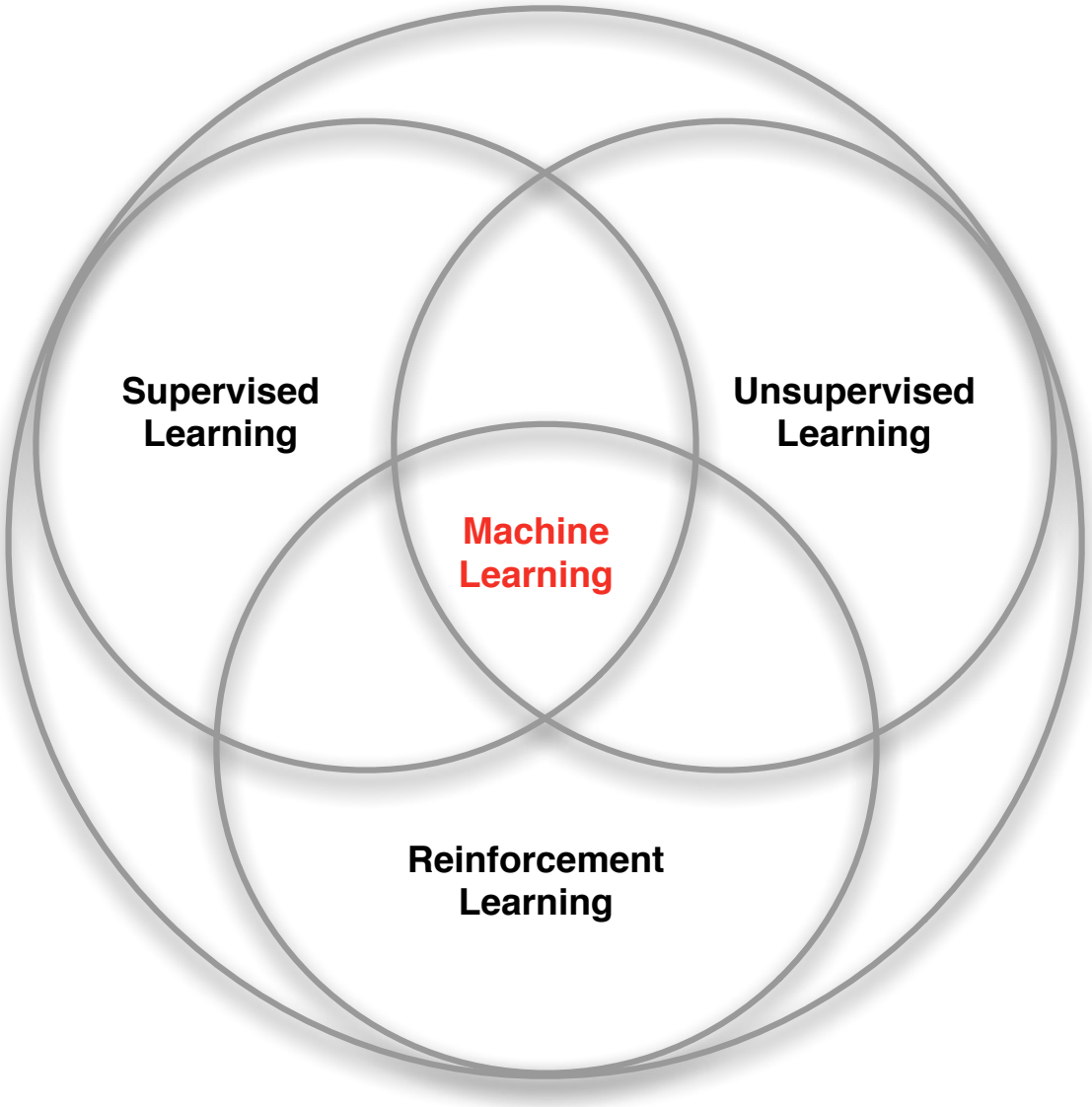
Learn to maximize the expected cumulative reward per episode

?

Reinforcement Learning vs Machine Learning

Reinforcement learning is one of three basic machine learning paradigms, alongside supervised learning and unsupervised learning.

Branches of Machine Learning



From David Silver's Slides

? Discussion ?

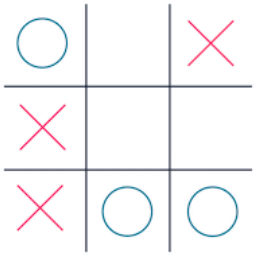
Topics for today

- What is reinforcement learning?
- Difference from other machine learning paradigms?
- Application stories.
- Topics to be covered in this course.
- Course logistics

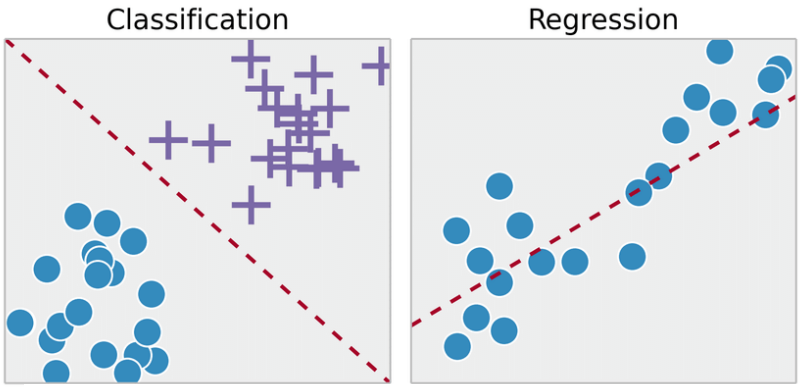
?

Other AI problems?

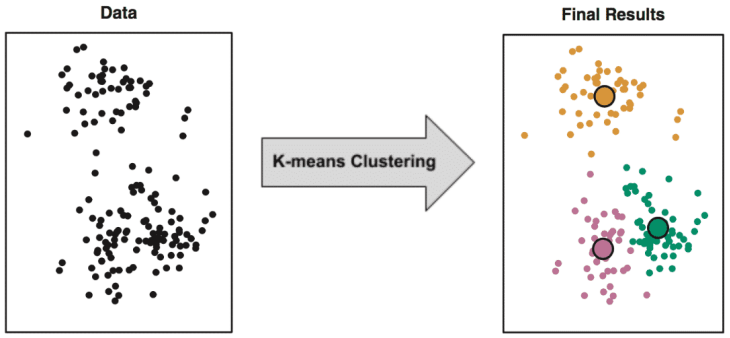
- AI Planning



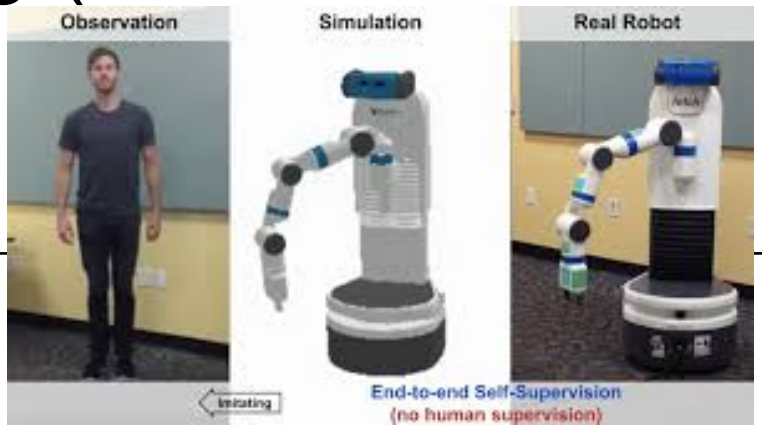
- Supervised learning



- Unsupervised learning



- Imitation learning (inverse reinforcement learning)



RL involves 4 key aspects

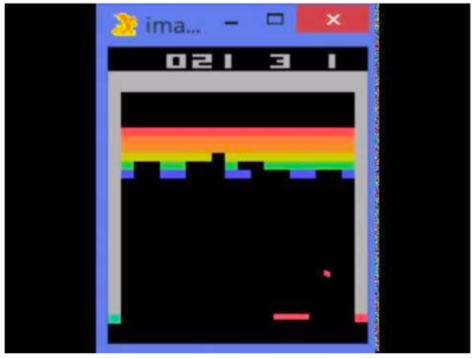
1. Optimization.

- Goal is to find an optimal way to make decisions, with maximized total cumulated rewards

2. Exploration.

2. Generalization.

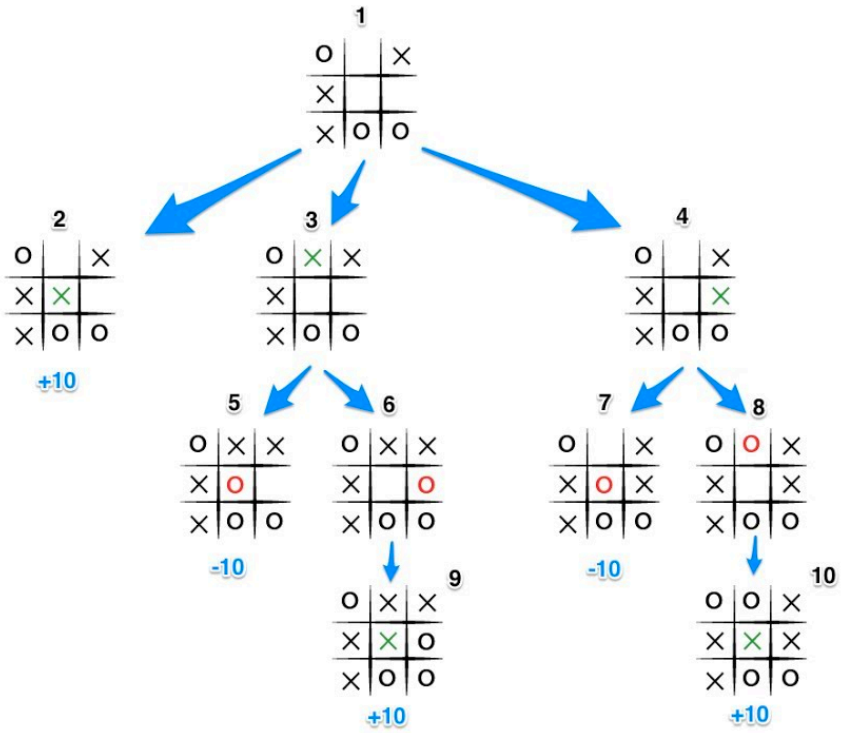
- Programming all possibilities is not possible.



4. Delayed consequences

AI planning vs RL

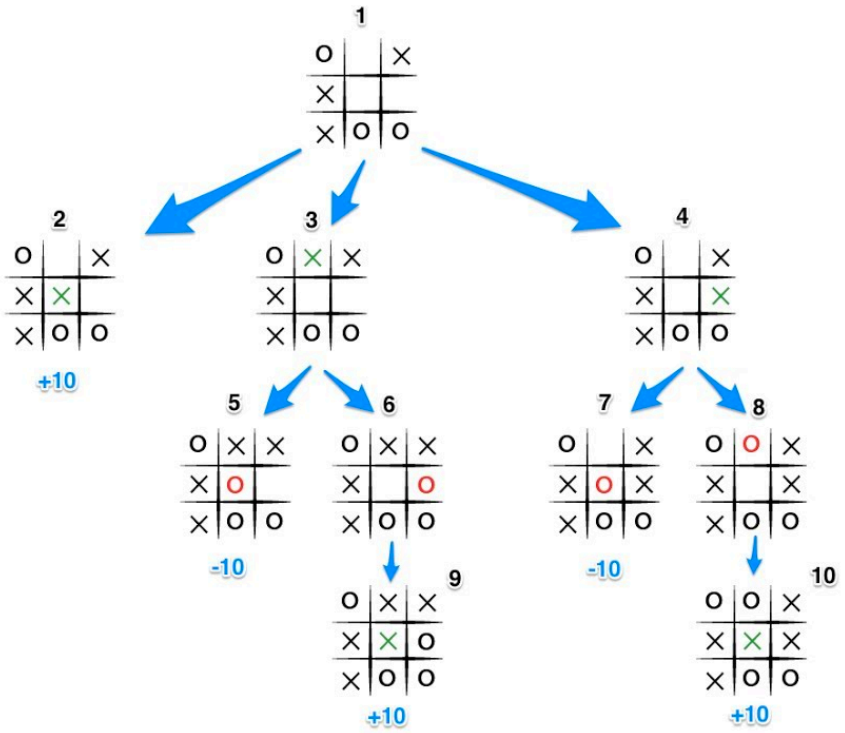
- AI planning: *?*
  - Optimization
  - Generalization
  - No Exploration
  - Delayed consequences



- Computes good sequence of decisions
- But given model of how decisions impact world

AI planning vs RL

- AI planning:
  - Optimization:
    - Objective: Reward
    - (e.g., likelihood of winning the game)
  - Generalization
    - Apply for all possible scenarios
  - No Exploration
  - Delayed consequences
    - A good move may lead to winning the game after multi-steps.

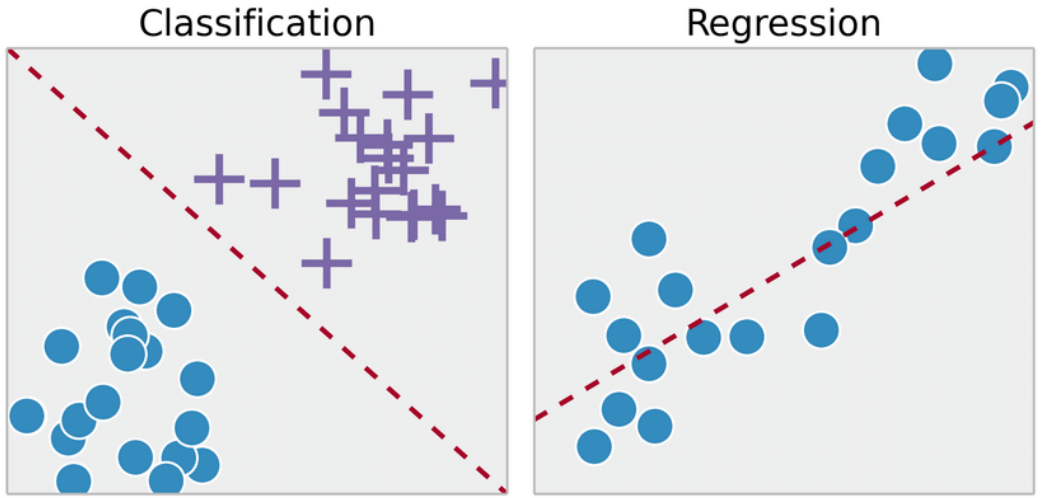


- Computes good sequence of decisions
- ~~But given model of how decisions impact world~~

?

Supervised Learning vs RL

- Supervised Learning: ?
  - Optimization
  - Generalization
  - No Exploration
  - No Delayed consequences

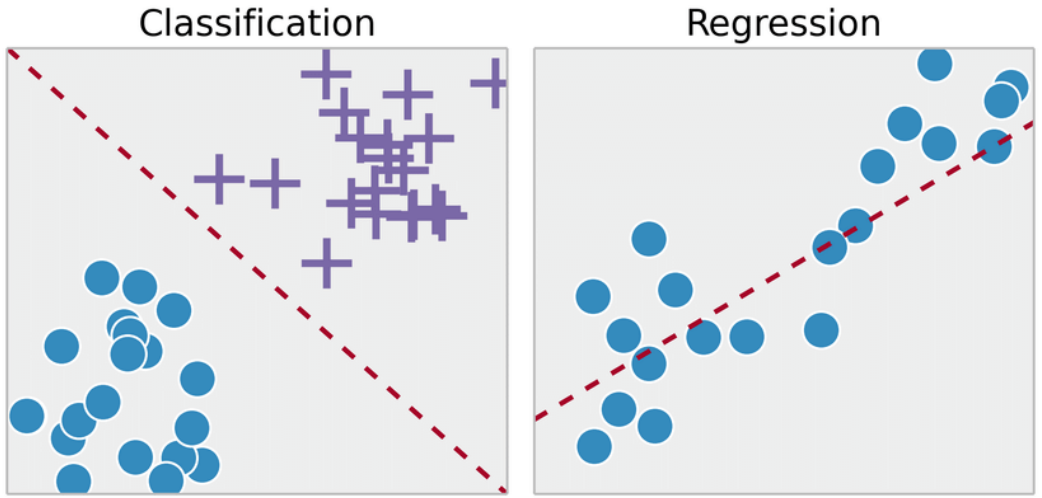


- Learns from experience
- But **provided** correct labels

?

Supervised Learning vs RL

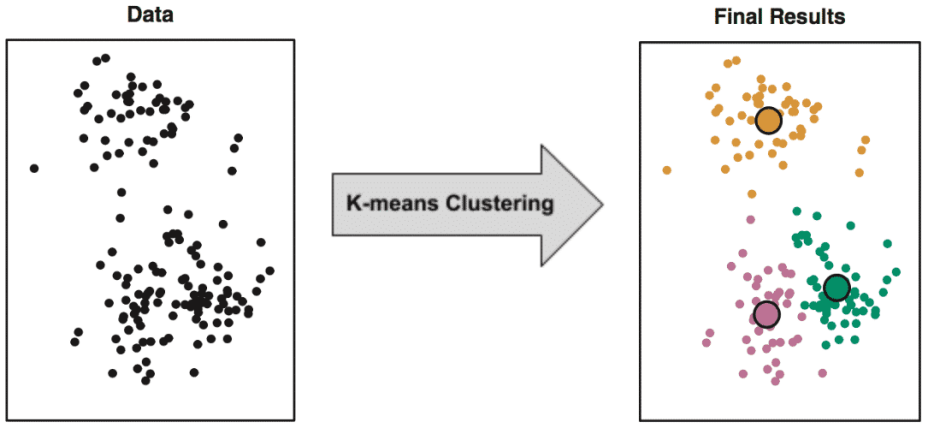
- Supervised Learning:
  - Optimization
    - Objective: Minimize the classification loss
  - Generalization
    - From training data to testing data
  - No Exploration
  - No Delayed consequences



- Learns from experience
- But **provided** correct labels

Unsupervised Learning vs RL

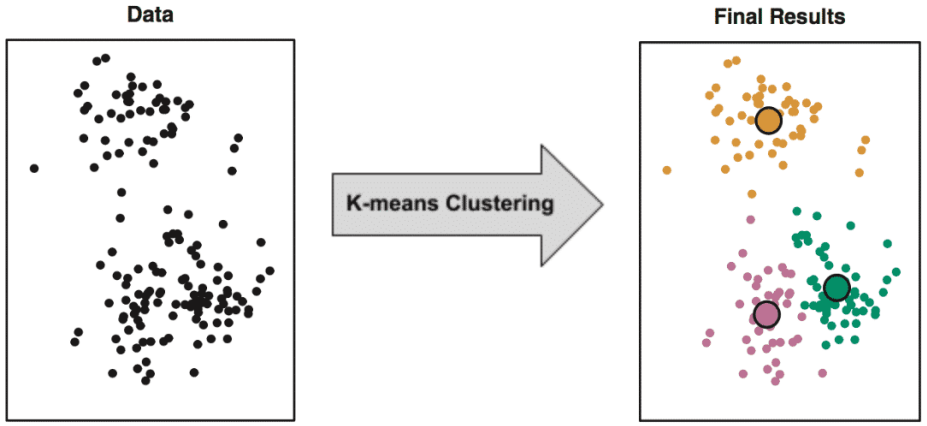
- Unsupervised Learning:?
  - Optimization
  - Generalization
  - No Exploration
  - No Delayed consequences



- Learns from experience
- But **no** labels from world

Unsupervised Learning vs RL

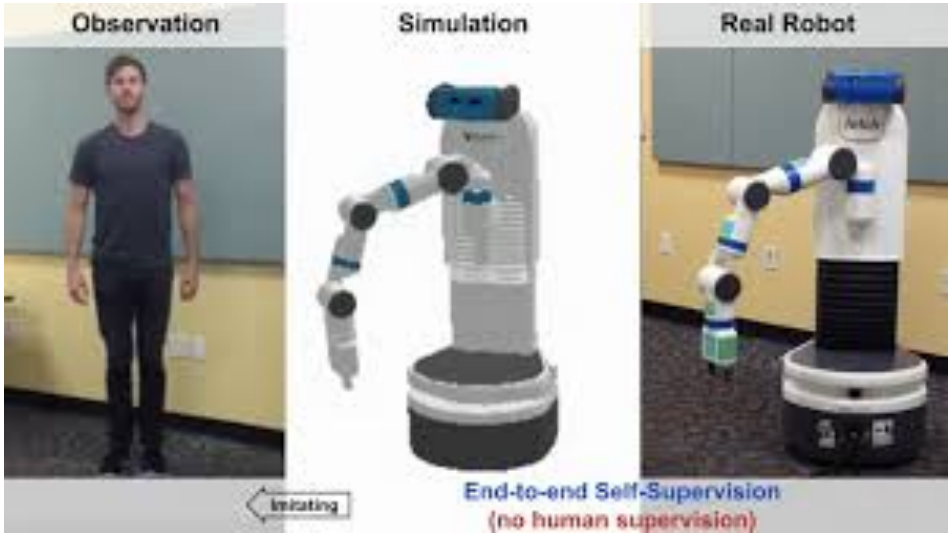
- Unsupervised Learning:
  - Optimization
    - e.g., k-means,
    - objective: minimize within-cluster distance
  - Generalization
    - e.g., k-means,
    - New data have the same clusters (centroids)
  - No Exploration
  - No Delayed consequences



- Learns from experience
- But **no** labels from world

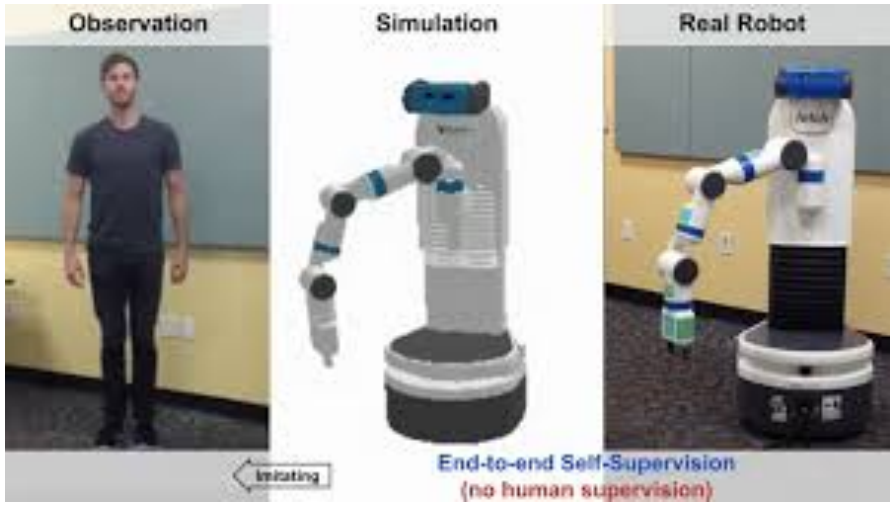
Imitation Learning vs RL

- Imitation Learning: *?*
  - Optimization
  - Generalization
  - No Exploration
  - Delayed consequences



- Learns from experience of others
- Assumes input demos of good policies

Imitation Learning vs RL



- Imitation Learning:

Given experts demonstration,

inversely infer experts' reward function.

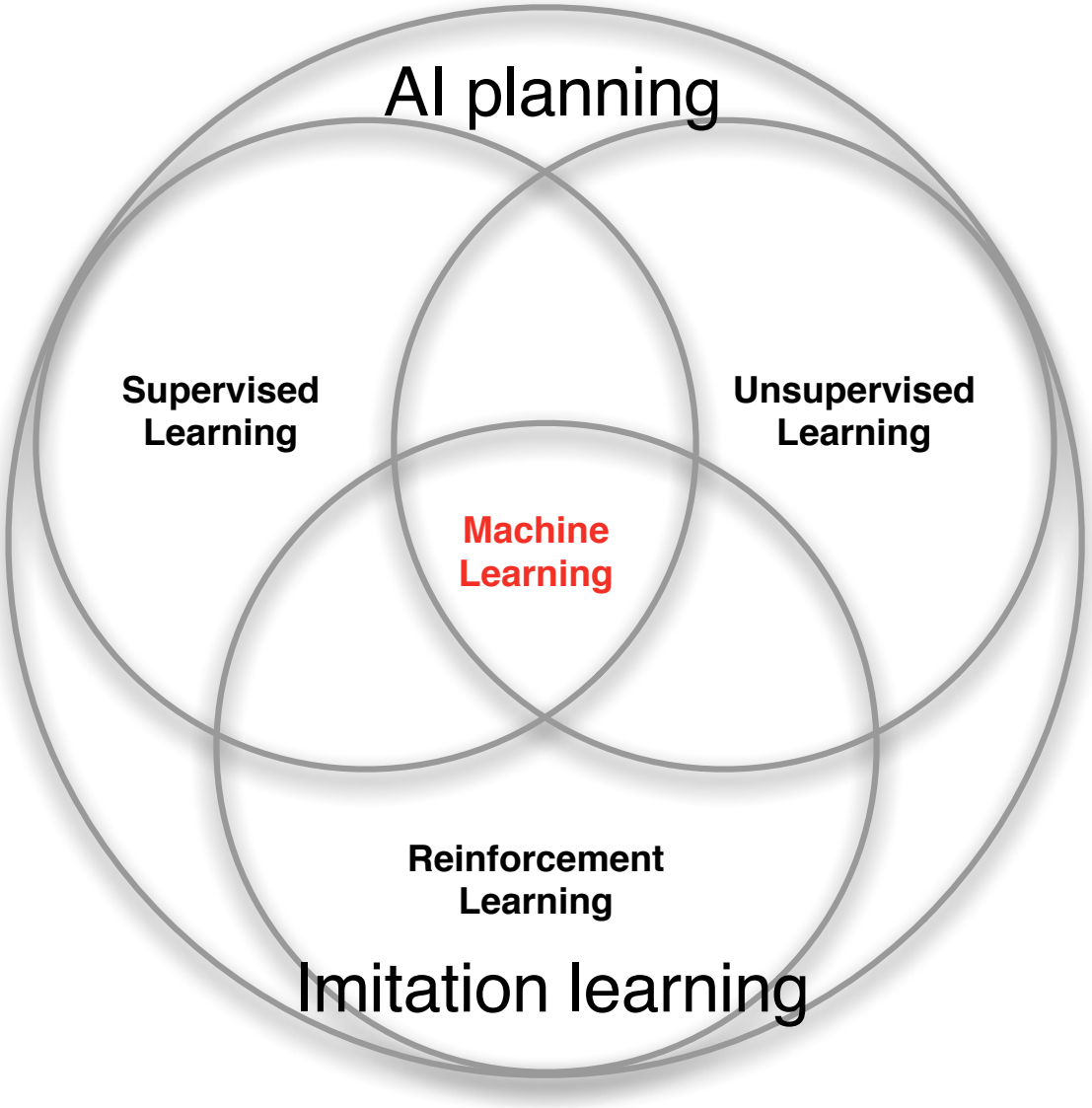
  - Optimization
    - –Objective: maximize the likelihood of the observed data
  - Generalization
    - –New data from the expert matches the learned reward function
  - No Exploration
  - Delayed consequences
    - –The same as RL

- Learns from experience of others
- ~~Assumes input demos of good policies~~

Reinforcement Learning

- Reinforcement Learning:
  - Optimization
    - Cumulative reward
  - Generalization
    - To all scenarios
  - Exploration
    - Evaluate the reward of different choices/actions
  - Delayed consequences
    - Sparse reward
- No data collected initially.
- Learning as collecting data through exploration

Branches of Machine Learning

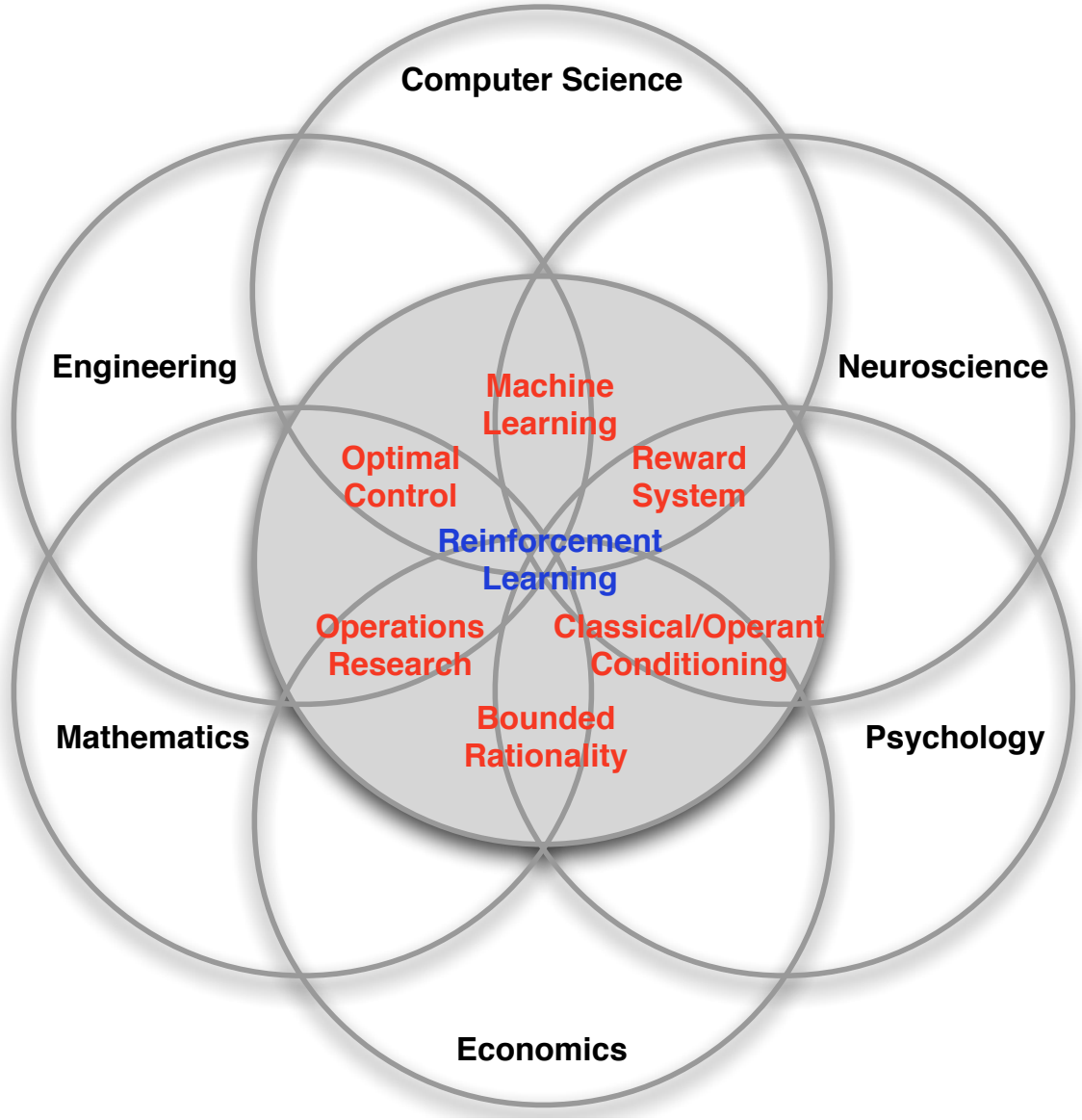


From David Silver's Slides

Topics for today

- What is reinforcement learning?
- Difference from Supervised and unsupervised machine learning?
- Application stories.
- Topics to be covered in this course.
- Course logistics

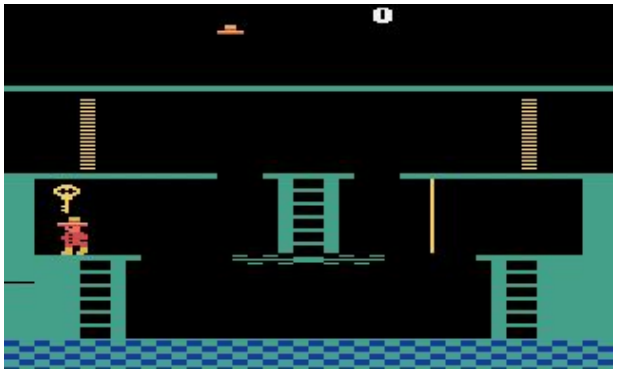
Many Faces of Reinforcement Learning



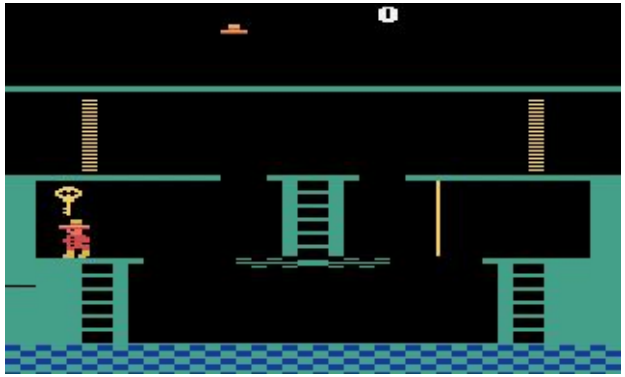
From David Silver's Slides

Why Now?

Amazing Reinforcement Learning Progress



Why Now?



≠

?

AI Challenges

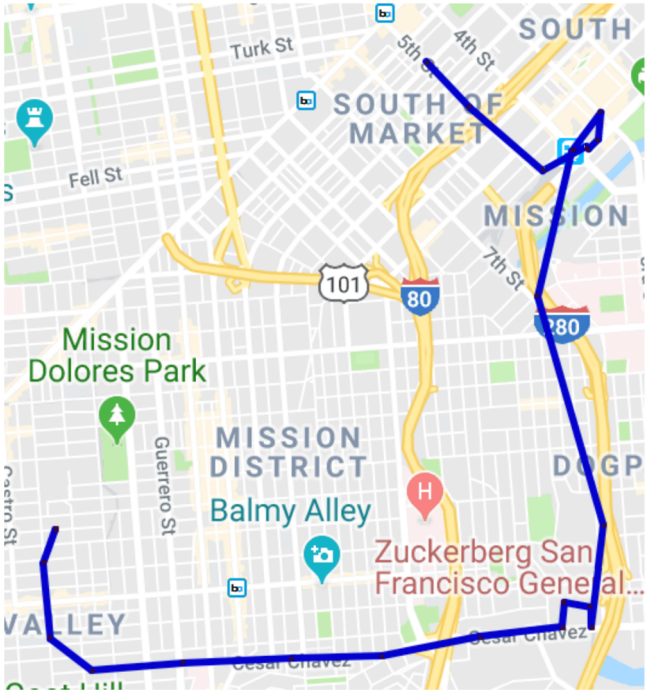
Research Story #1

- Package delivery system planning
  - Multi-agents
  - Coorporative game
  - Multi-agent RL

- Efficient and Effective Express via Contextual Cooperative Reinforcement Learning, Yexin Li (The Hong Kong University of Science and Technology);Yu Zheng (Urban Computing Business Unit, JD Finance);Qiang Yang (The Hong Kong University of Science and Technology), KDD 2019;

Research Story #2

- Outlier detection on trajectory data
  - Learn reward function of normal drivers
    - With inverse RL
  - Detect malicious drivers if
    - The reward function is
    - Significantly different
    - From the normal drivers



Anomalous taxi route

- Sequential Anomaly Detection using Inverse Reinforcement Learning, Min-Hwan Oh (Columbia University);Garud Iyengar (Columbia University); KDD 2019;

Research Story #3

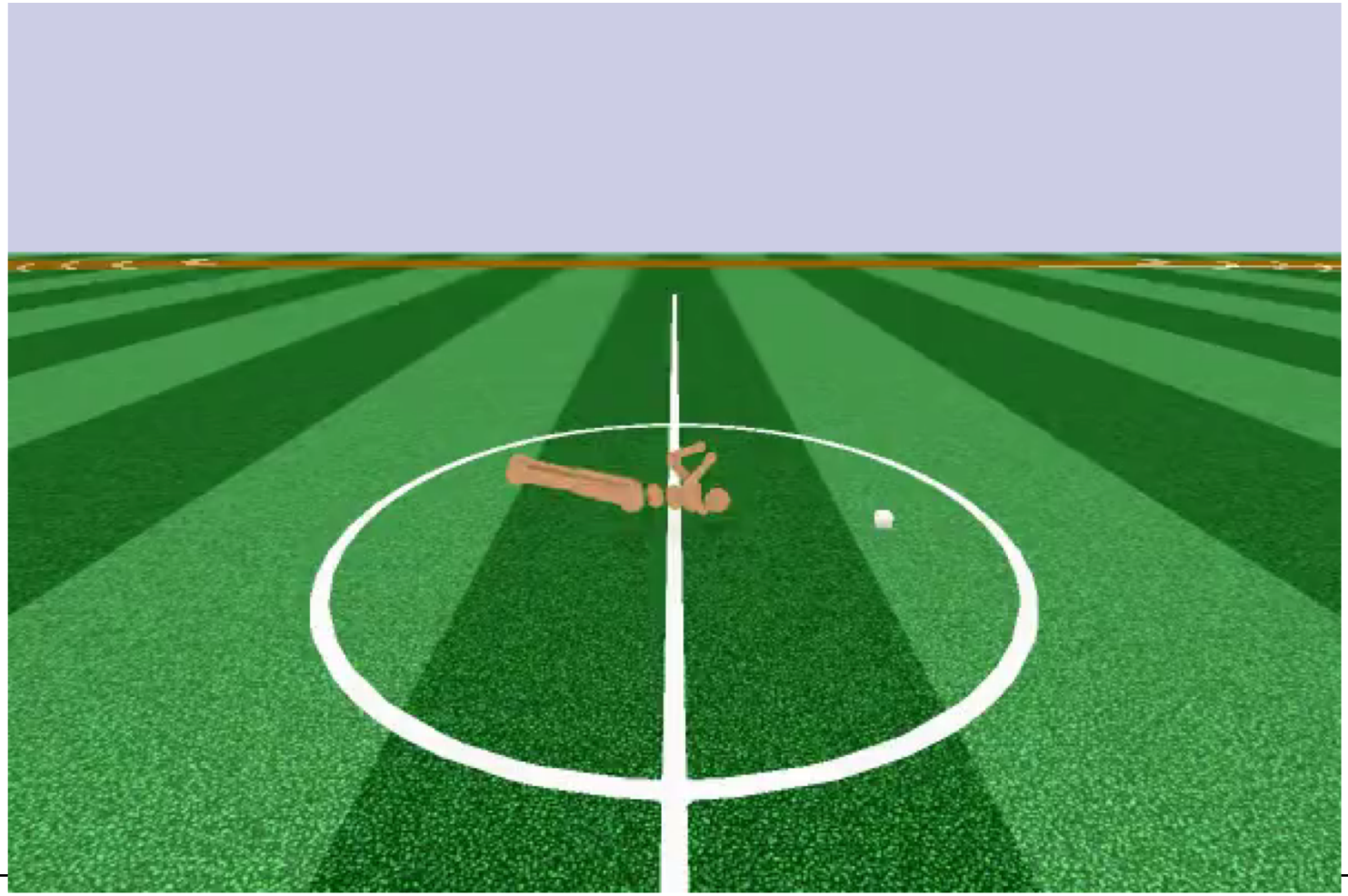
- Recommender system
  - Recommend product, news, photo feeds to keep long-term user engagement

- Reinforcement Learning to Optimize Long-term User Engagement in Recommender Systems, Lixin Zou, Long Xia, Zhuoye Ding, Song Jiaxing, Weidong Liu and Dawei Yin; KDD 2019;

Research Story #4 DeepMind

https://youtu.be/gn4nRCC9TwQ

Research Story #5 OpenAI

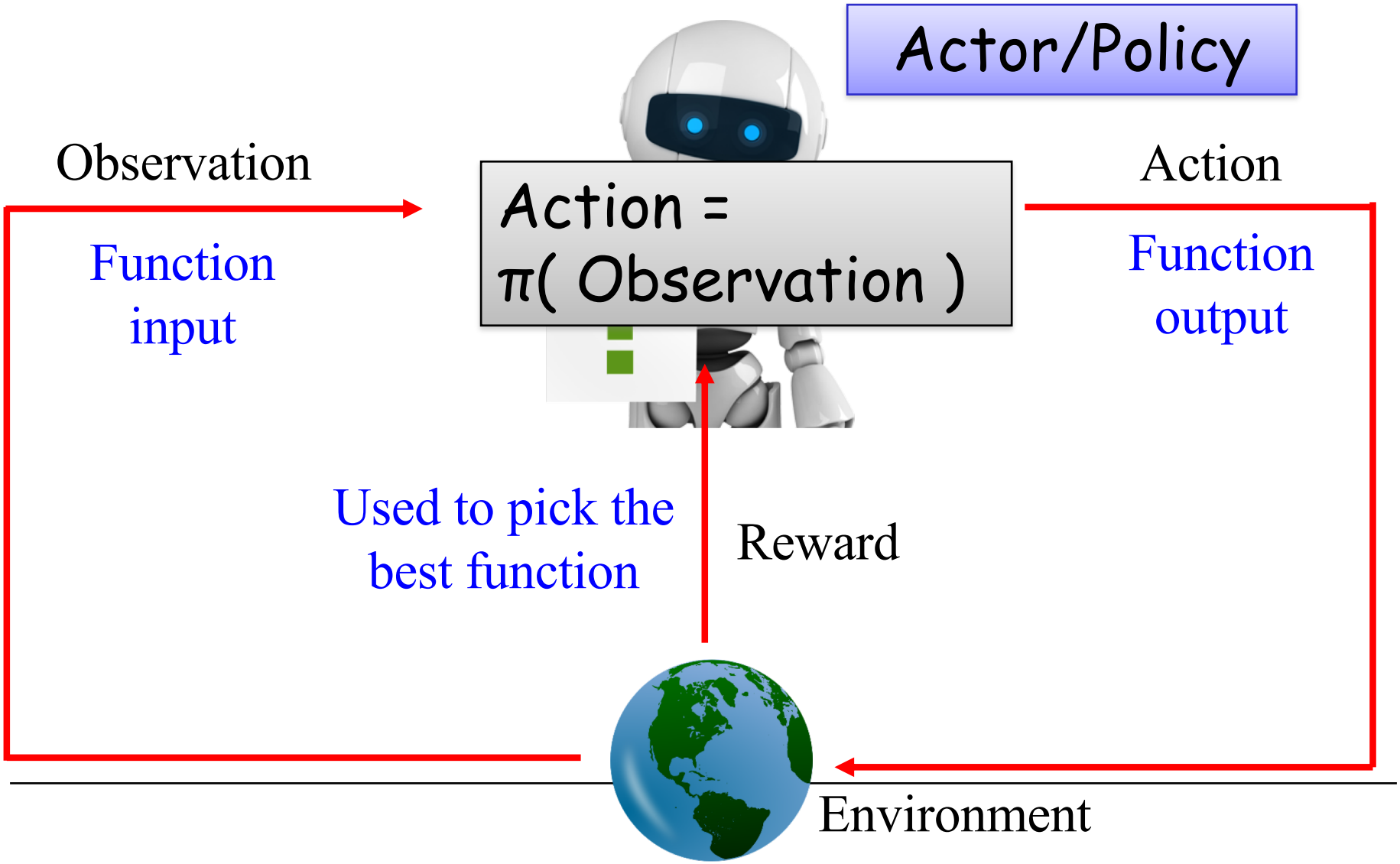


https://blog.openai.com/openai-baselines-ppo/

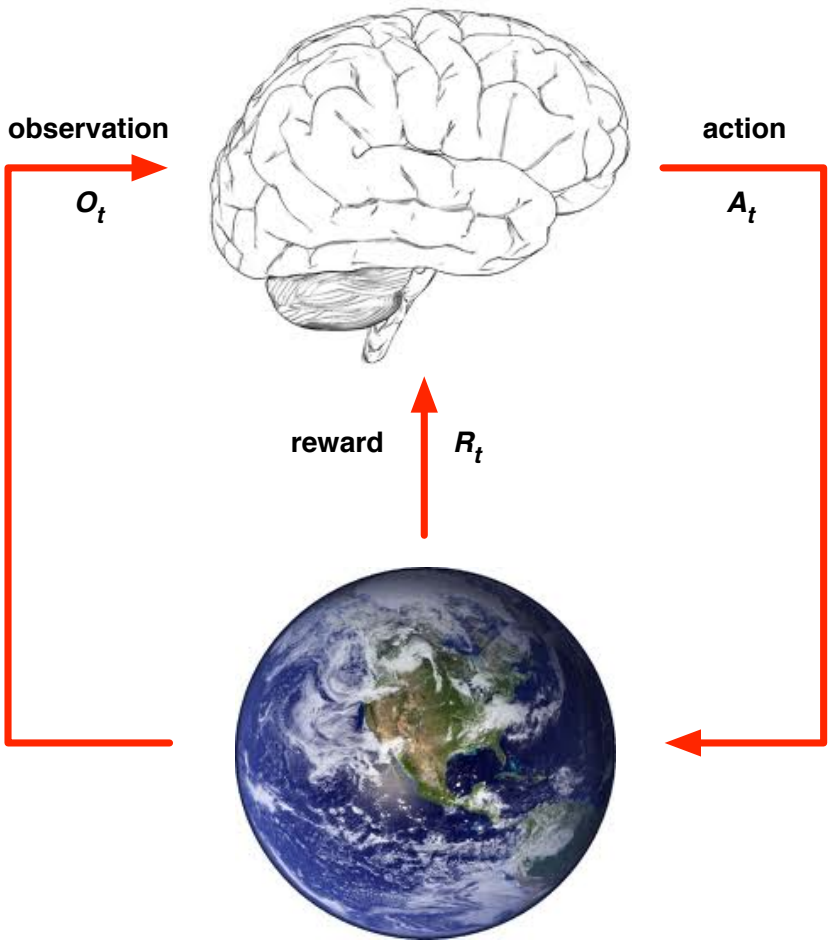
Topics for today

- What is reinforcement learning?
- Difference from Supervised and unsupervised machine learning?
- Application stories.
- Topics to be covered in this course
- Course logistics

Reinforcement Learning



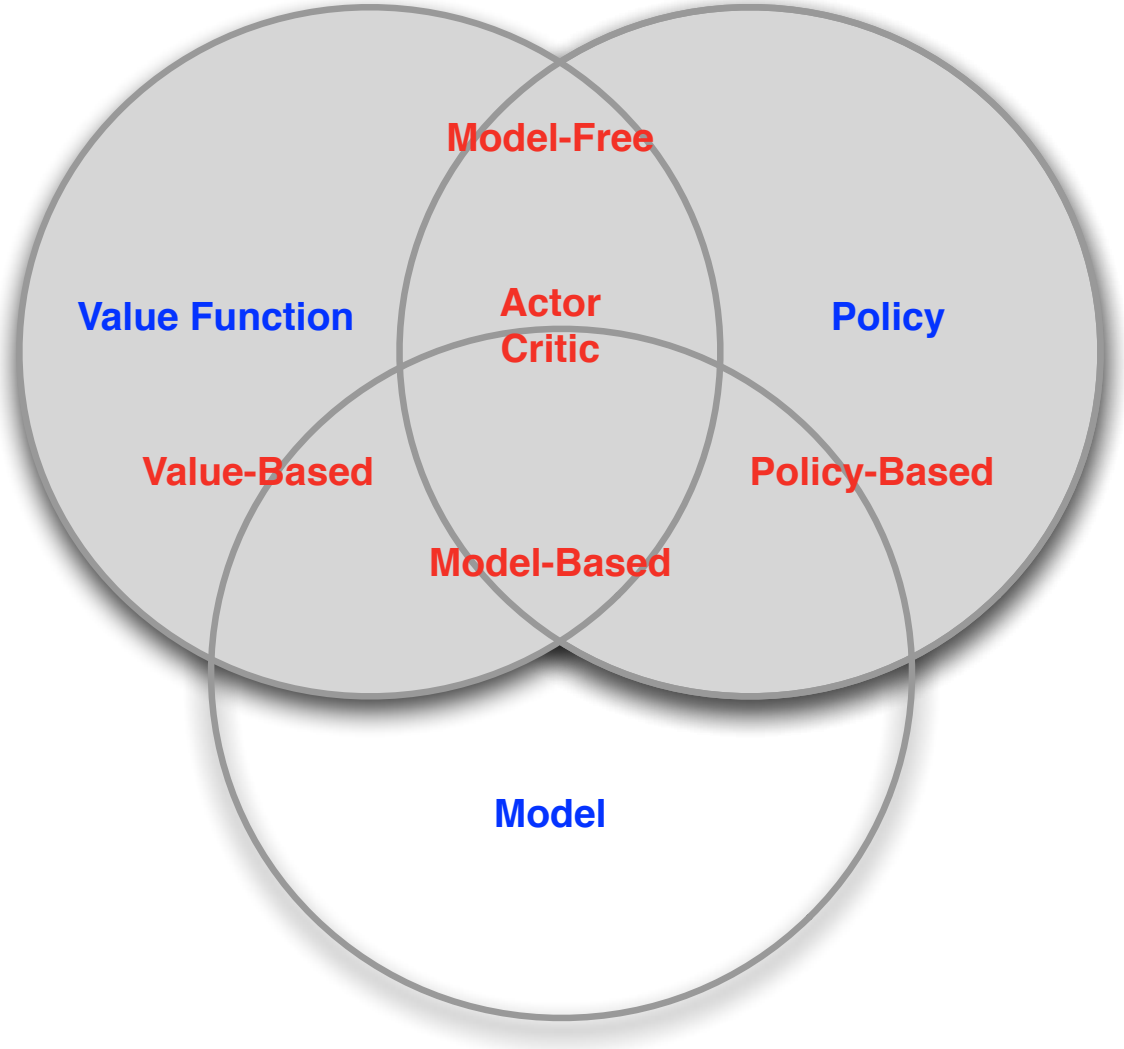
Agent and Environment



- At each step t the agent:
  - Executes action A_t
  - Receives observation O_t
  - Receives scalar reward R_t
- The environment:
  - Receives action A_t
  - Emits observation O_{t+1}
  - Emits scalar reward R_{t+1}
- t increments at env. step

From David Silver's Slides

RL Agent Taxonomy



From David Silver's Slides

RL Topics You will Learn

- Reward in Tabular representation
  - Model-based Planning, Policy Evaluation, and Control
  - Model-free Policy Evaluation, and Control
    - Monte Carlo, Temporal difference, SARSA, Q-Learning
- Reward as a function representation
  - Linear function: Approximation and Control
  - Non-linear function (Deep reinforcement learning) (Review DL)
    - DQN (Deep Q-Learning), Policy Gradient, PPO, TPRO
- Imitation Learning (Inverse RL)
  - Linear reward function
  - Non-linear reward function
    - Solution with Generative Adversarial Network (GAN) (Review GAN)
- Applications/Extensions:
  - ~~Sequence Generation (e.g., Sentence generation)~~
  - Relation to Auto-Encoder,
  - Meta-RL, Multi-Agent RL, Adversarial Attack to RL/IRL, etc.

Statistics

1. DS/CS/RBE
2. 2nd+ year Graduate
3. DS/CS/RBE 2+nd year
4. PhD

Course Prerequisite

- Pre-requisites
- 1. Python proficiency
  2. Basic probability and statistics, Multivariate calculus and linear algebra
  3. Machine learning or AI (e.g. CS229, CS221)
  4. Deep learning
  5. Fine if you don't know DL before taking RL.
  - We will cover the basics, but quickly.

More importantly

- Willing to learn and work hard
- Love to ask questions and solve problems

Course Mechanisms

- A lecture- and project-oriented course
- A series of lectures combining both theory and Practices in two "parallel" tracks:
  - Track 1: lectures
    - Foundations of RL, with 5 quizzes
  - Track 2: Projects
    - 3 individual projects
    - 1 Group project

Course Materials

- Textbooks
  - No Textbook.
  - Recommendation: Reinforcement Learning: An Introduction, Sutton and Barto, 2nd Edition. This is available for free here and references will refer to the final pdf version available here.
- Assigned readings with each class:
  - Reading materials on class website (tentatively, updated as we go along)
  - Optional papers for background, supplementary and further readings
- Slides
  - Will be posted on the class website before each class

Course Requirements

- Do assigned readings
  - Be prepared, read and review required readings *on your own in advance!*
- Complete course projects
  - Both individual and group projects
- Attend and participate in class activities
  - Please ask and answer questions in (and out of) class!
  - Let's try to make the class interactive and fun!

Class Information

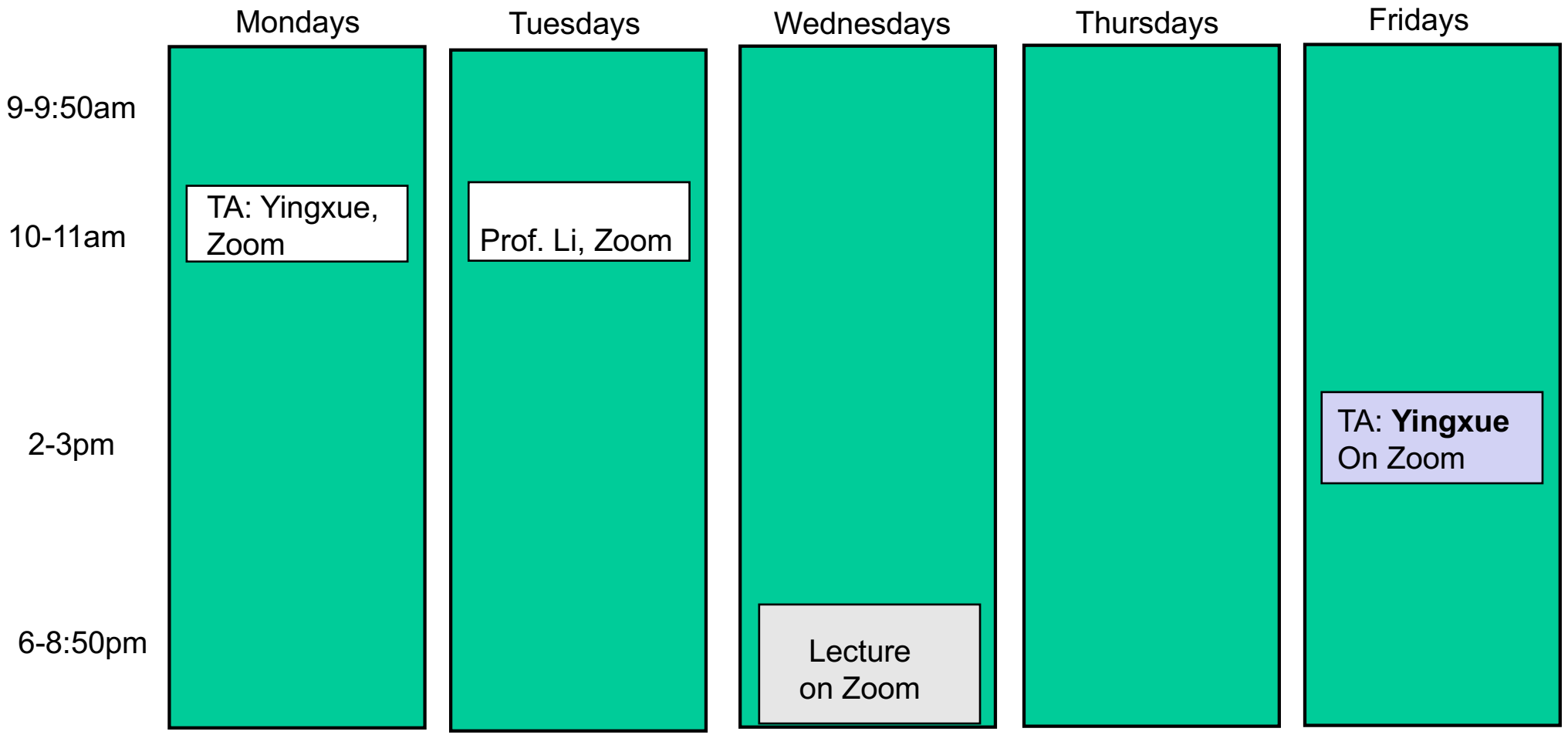
- Class Website :
  - https://users.wpi.edu/~yli15/courses/DS595Spring22/index.html
- Announcement Page
  - Check the Canvas/Email periodically
- Email address Q&As, discussions, etc.
  - Professor: yli15@wpi.edu
  - TA：yzhang31@wpi.edu

Office Hours

- Professor Li's Office Hours:
  - Zoom (Link is available on Canvas)
  - Email: yli15@wpi.edu
  - Tues 10-11AM
  - Others by appointments
- TA Yingxue Zhan's Office Hours:
  - Zoom (Link is available on Canvas)
  - Email: yzhang31@wpi.edu
  - Mon 10-11AM, & Fri 2-3PM
  - Others by appointments

Class Calendar & Office Hours

All sessions are online.



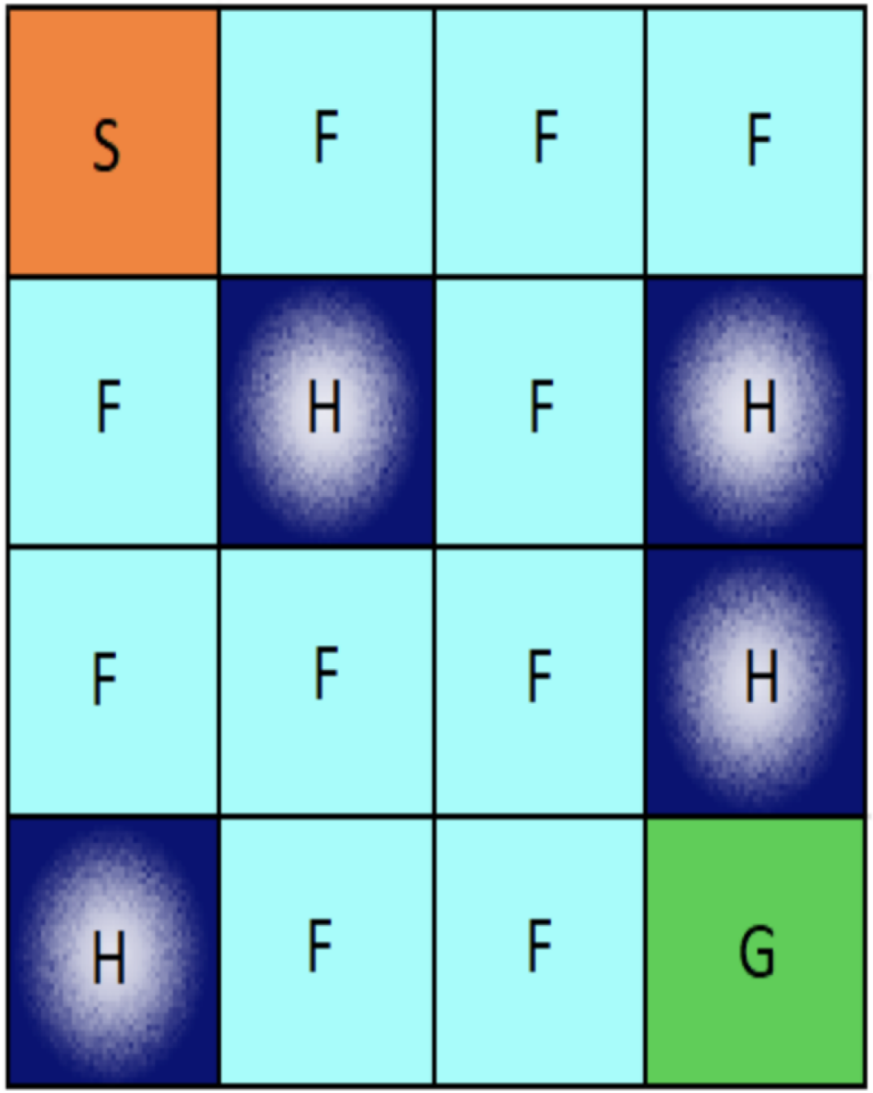
Workload and Grading

- Workload
  - Oral work (10%)
  - Quizzes, Exams (30%): 5 quizzes
  - Projects (60%);
    - Project 1 for 5%,
    - Project 2 for 10%,
    - Project 3 for 15%,
    - Project 4 for 30%)

- Focus more on critical thinking, problem solving, “heads-on/hands-on” experience!
  - Understand, formulate and solve problems
  - 4 Course Projects

Project 1 (Model-based Planning)

- Implement Dynamic Programming
- Play with OpenAI Gym (Frozen Lake)
  - The agent moves through a 4*4 gridworld
  - The agent has 4 potential actions:
    - LEFT = 0
    - DOWN = 1
    - RIGHT = 2
    - UP = 3
  - The action is stochastic:
    - Stochastic: the action may move to several states based on transition probability.
    - Deterministic: the action will only move to one state.

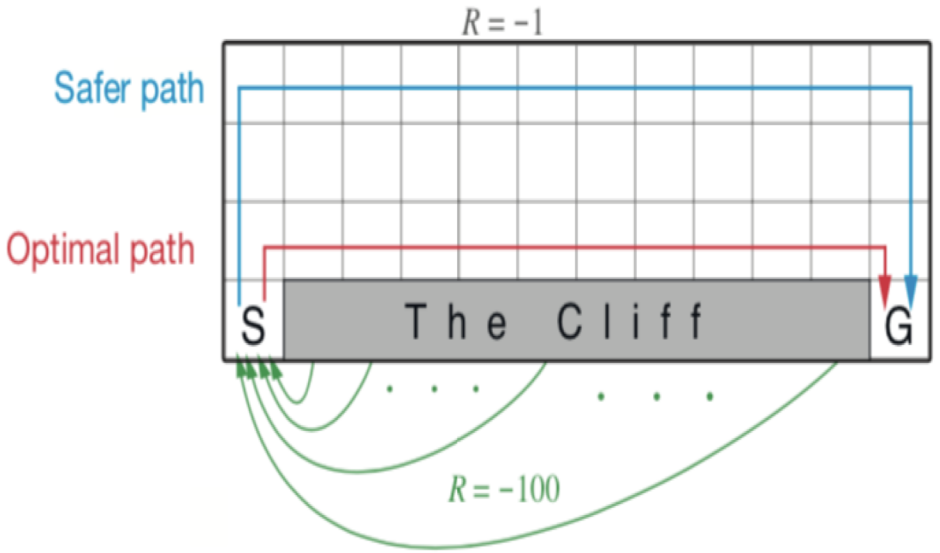


Project 2-1 MC (model-free)

- Implement Monte Carlo
- Play with OpenAI Gym (BlackJack)
  - Obtain cards the sum of whose numerical values is as great as possible without exceeding 21
  - Each state is a 3-tuple of:
    - The player's current sum
    - The dealer's face up card
    - Whether or not the player has a usable ace
  - The agent has two potential actions:
    - STICK = 0
    - HIT = 1

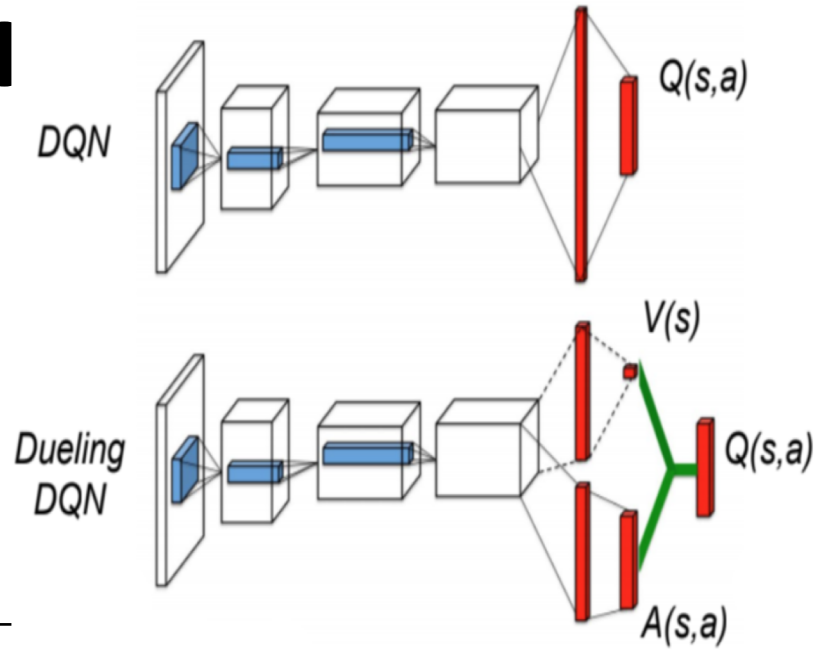
Project 2-2 TD (model-free)

- Implement SARSA(on-policy) and Q-Learning(off-policy)
- Play with OpenAI Gym (CliffWalking)
  - The agent moves through a 4*12 gridworld
  - The agent has 4 potential actions:
    - LEFT = 0
    - DOWN = 1
    - RIGHT = 2
    - UP = 3



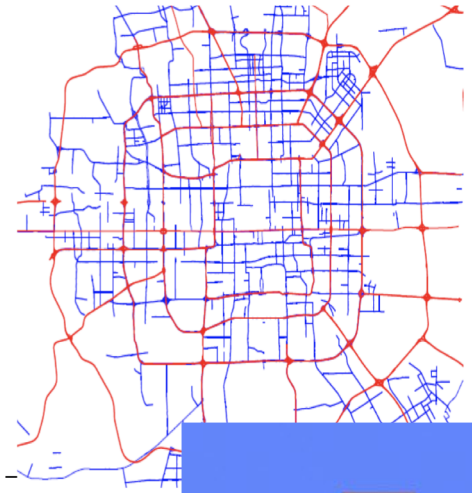
Project 3 DQN

- Implement Deep Q-Learning
- Play with OpenAI Gym (Breakout)



Project 4

Explore RL in more depth!



Example: MineRL competition: Minecraft Obtain Diamond task. Jun-Oct. 2019. http://minerl.io/competition/

Course Project 4

- Projects will be in groups!
  - Around 4 students per group, depending on enrollment

- "research-oriented" project timeline: (tentative!)
  - Team Project

  Week 10 (3/23 W), Starting date
  Week 11 (3/30 W), Proposal Due. (Upload it to Canvas)
  Week 13 (4/13 W), Progressive report due (Upload it to Canvas discussion board)
  Week 15 (4/25 M), Project report due. (Upload it to Canvas discussion board)
  Week 15 (4/27 W), Project poster session. (On Zoom)

Class Resources

- Presentation
  - https://users.wpi.edu/~yli15/courses/DS595Spring22/Presentation.html
- More resources
  - http://users.wpi.edu/~yli15/courses/DS595Spring22/Resources.html

Questions?