

# Reinforcement Learning

CMPT 729

Jason Peng

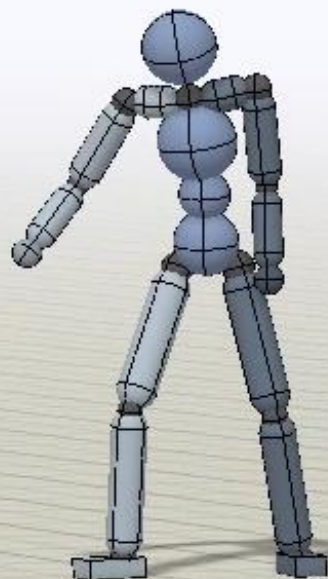
# Overview

---

- What is reinforcement learning?
- Applications
- Logistics

# About me

---



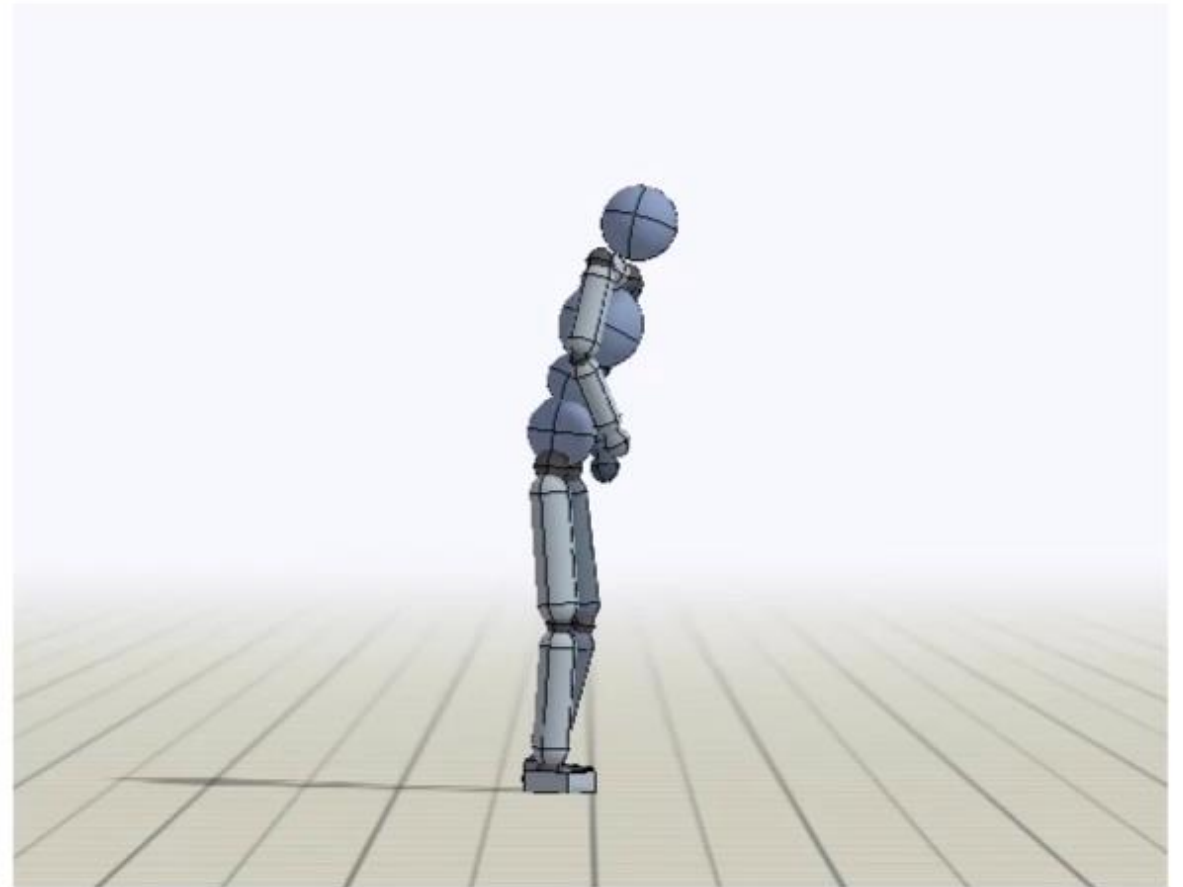
**DeepMimic: Example-Guided Deep Reinforcement Learning of Physics-Based Character Skills**  
Xue Bin Peng, Pieter Abbeel, Sergey Levine, Michiel van de Panne  
SIGGRAPH 2018

# About me

---



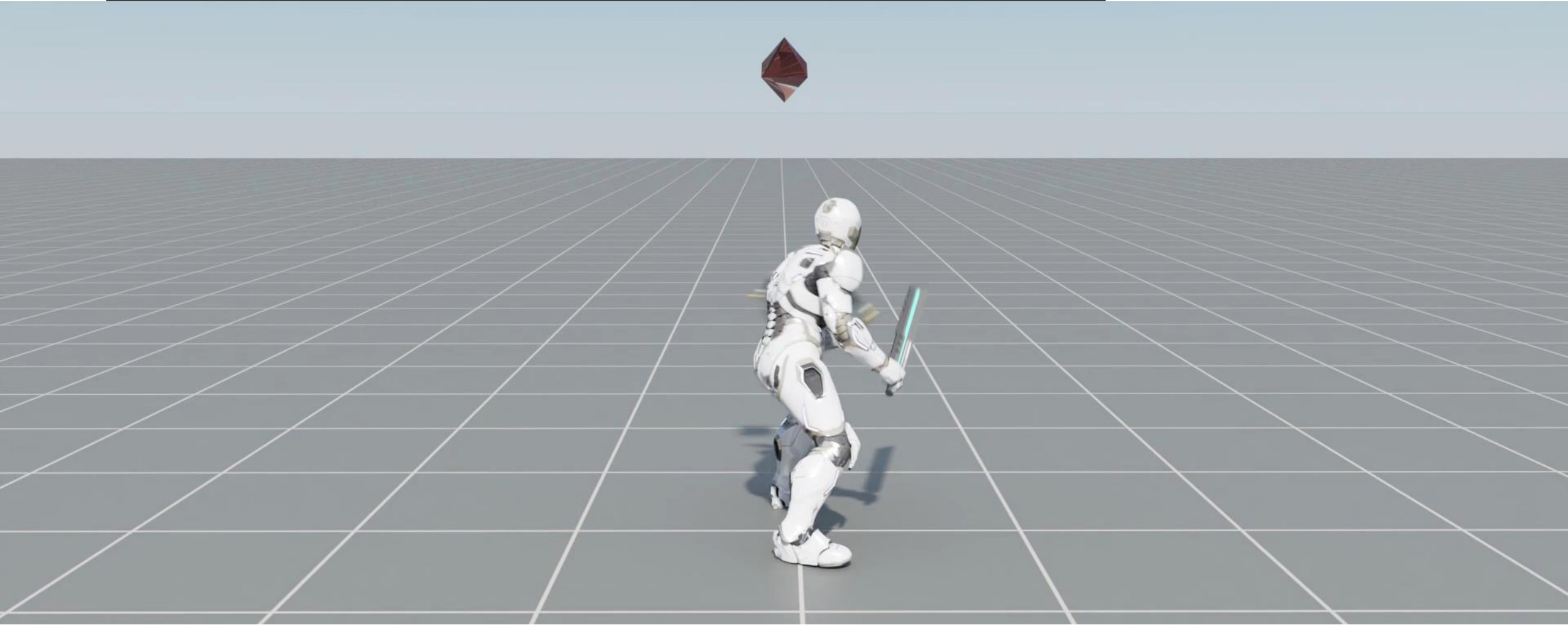
Video: Backflip B



Policy

# About me

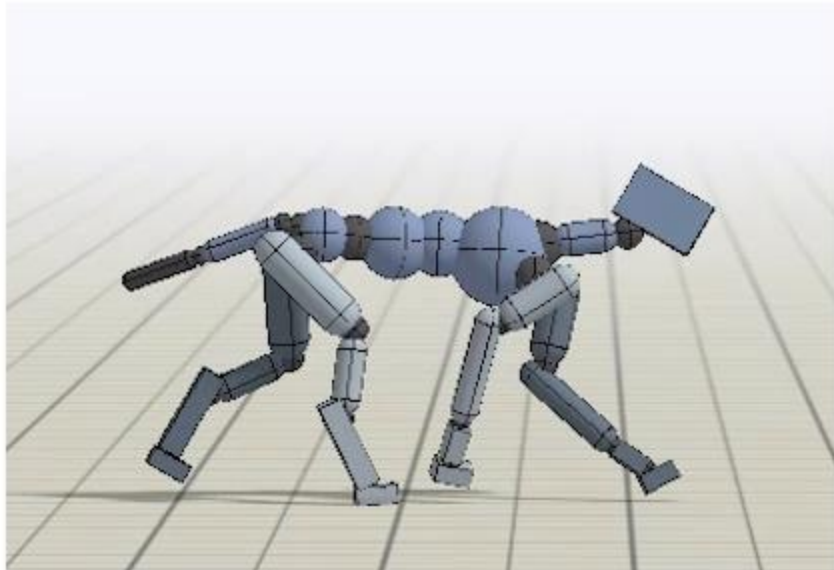
---



**ASE: Large-Scale Reusable Adversarial Skill Embeddings for Physically Simulated Characters**  
Xue Bin Peng, Yunrong Guo, Lina Halper, Sergey Levine, Sanja Fidler  
SIGGRAPH 2022

# About me

---



Reference



Simulation



Real Robot

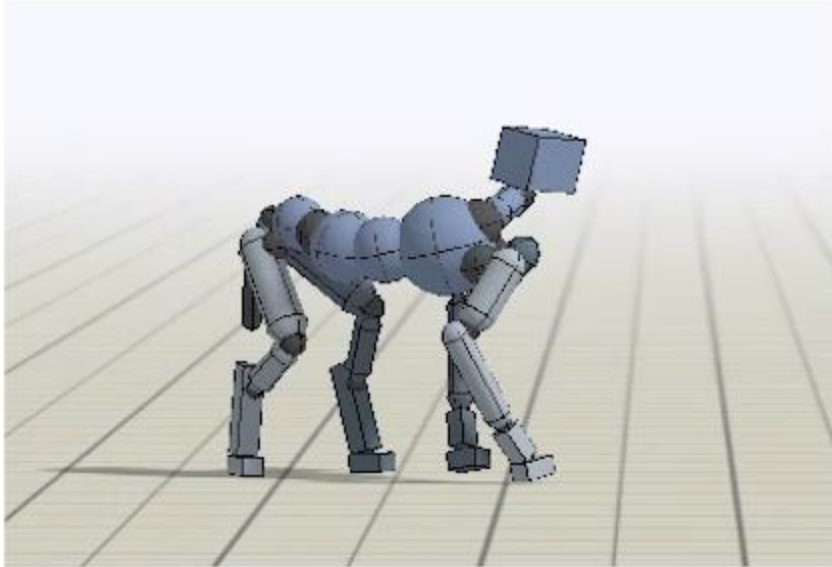
**Learning Agile Robotic Locomotion Skills by Imitating Animals**

Xue Bin Peng, Erwin Coumans, Tingnan Zhang, Tsang-Wei Edward Lee, Jie Tan, Sergey Levine

RSS 2020

# About me

---



Reference



Simulation



Real Robot

**Learning Agile Robotic Locomotion Skills by Imitating Animals**

Xue Bin Peng, Erwin Coumans, Tingnan Zhang, Tsang-Wei Edward Lee, Jie Tan, Sergey Levine

RSS 2020

What is **Reinforcement Learning**?



# What is Reinforcement Learning

---

**Reinforcement Learning** = Area of machine learning that studies techniques for solving **decision making** problems.

# Decision Making Problems

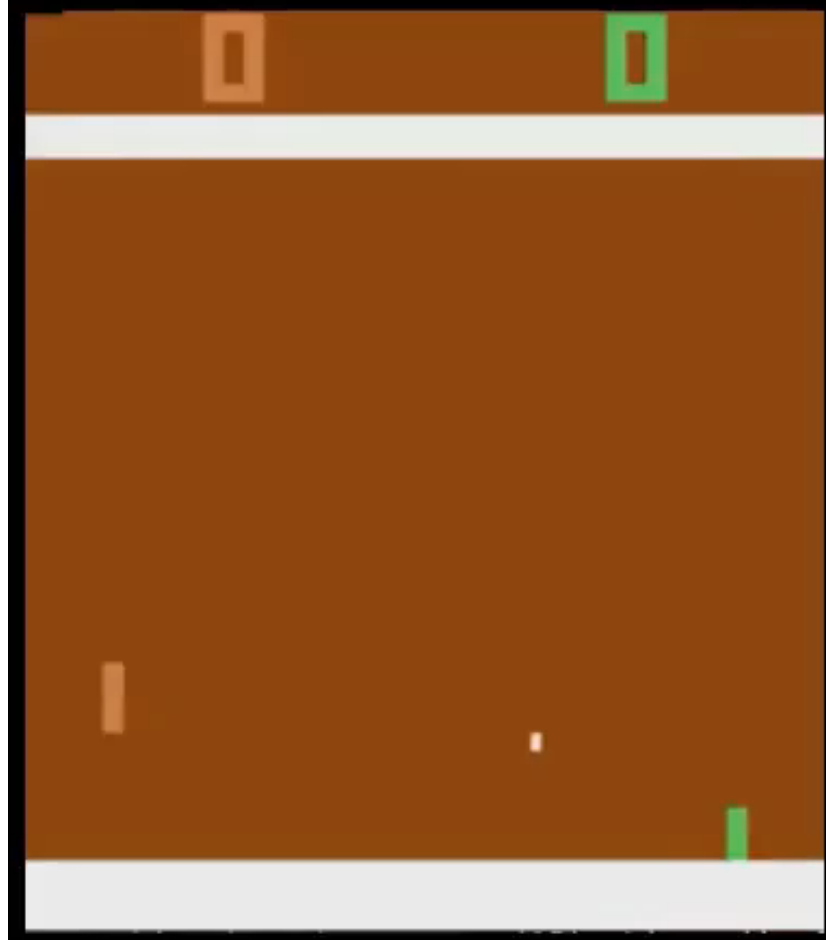
---



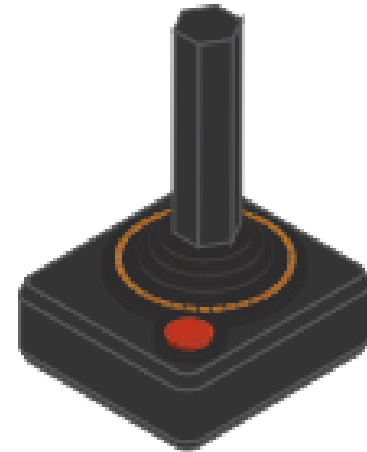
[Garry Kasparov vs. Deep Blue 1997]

# Decision Making Problems

---

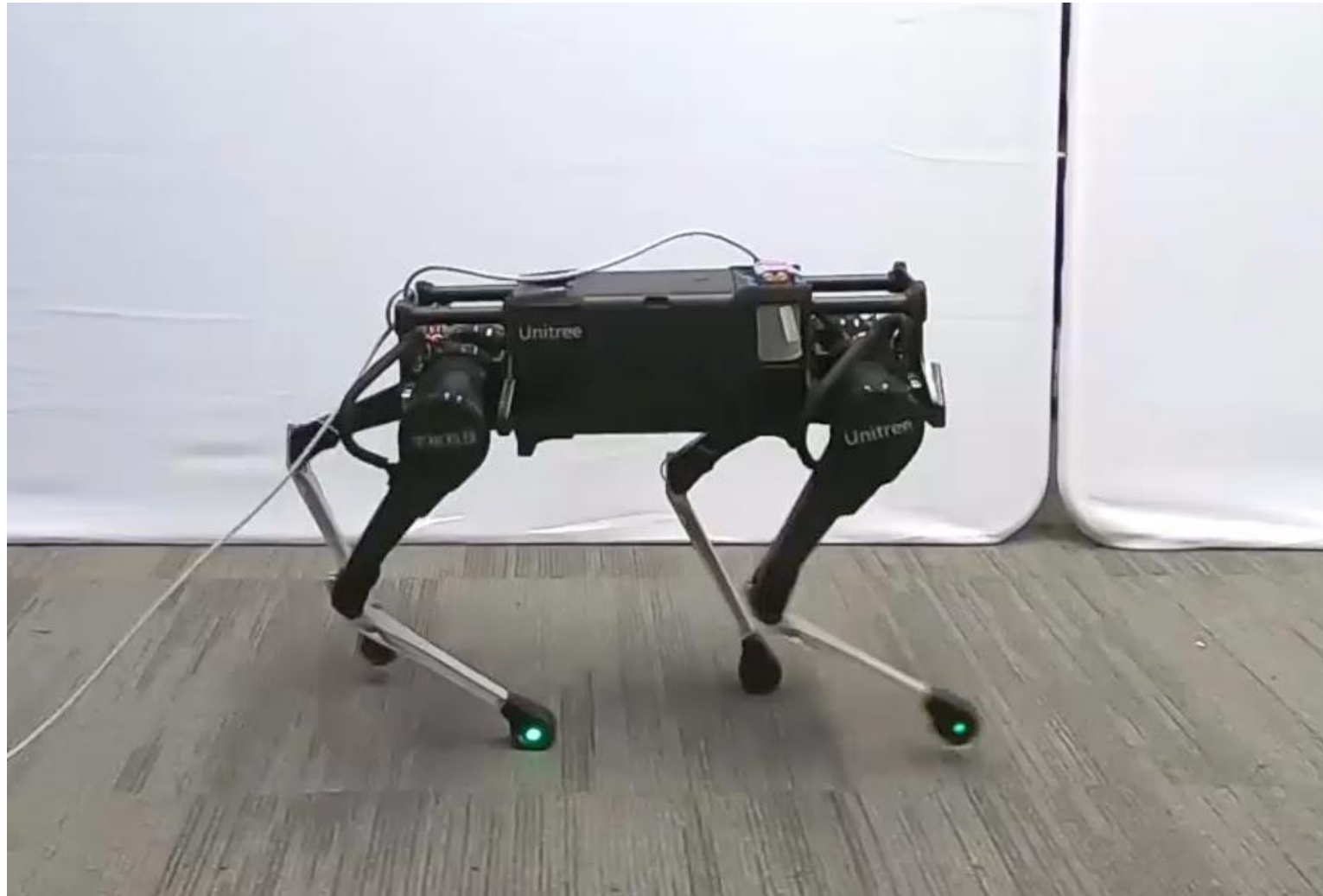


[Pong]

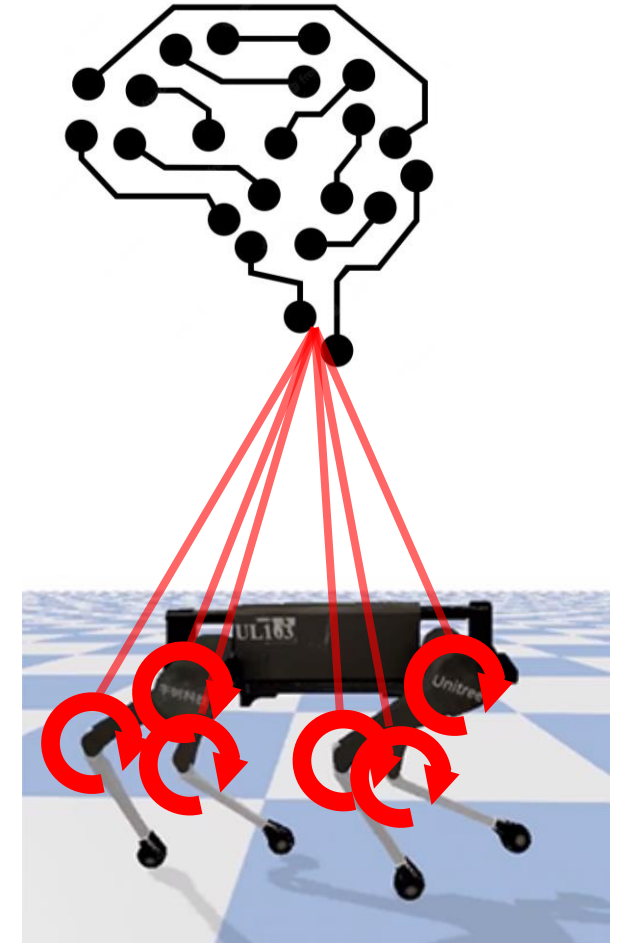


# Decision Making Problems

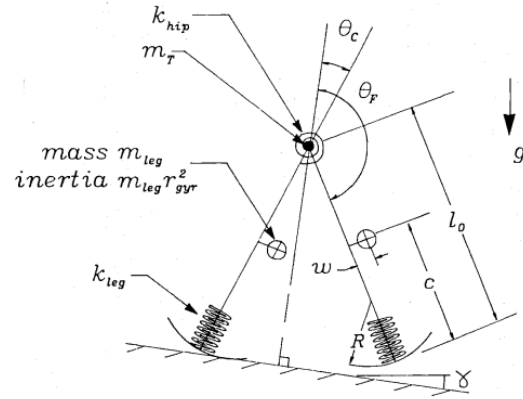
---



Controller



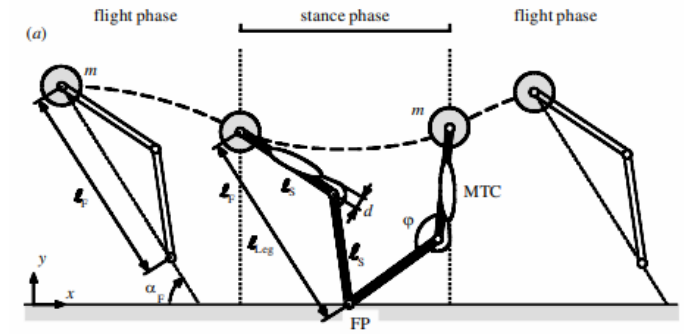
# Manual Controller Design



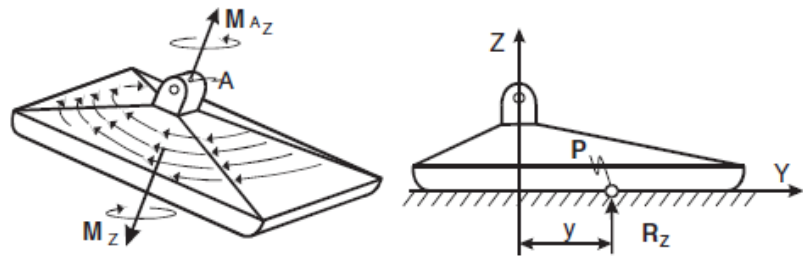
[McGeer 1990]



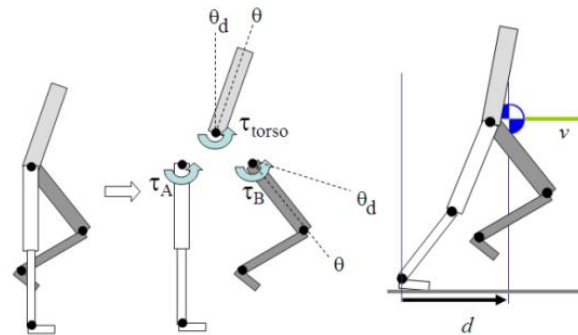
[Raibert and Hodgins 1991]



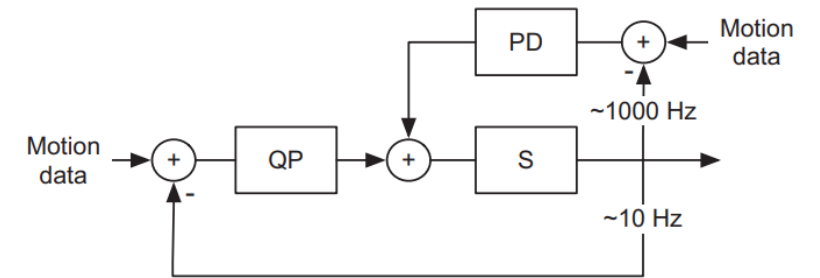
[Geyer et al. 2003]



[Vukobratović and Borovac 2004]



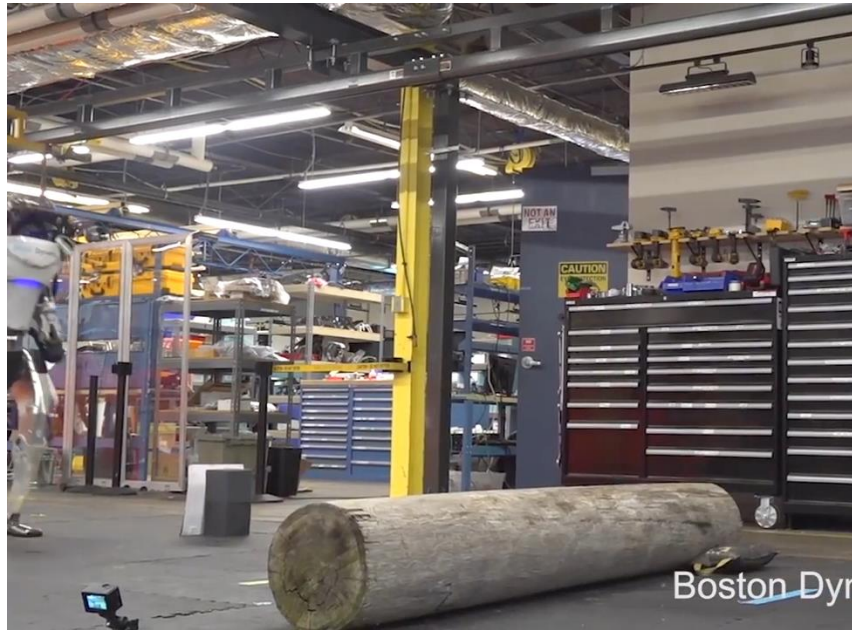
[Yin et al. 2007]



[Da Silva et al. 2008]

# Manual Controller Design

---



[Boston Dynamics 2018]



[ANYbotics 2018]

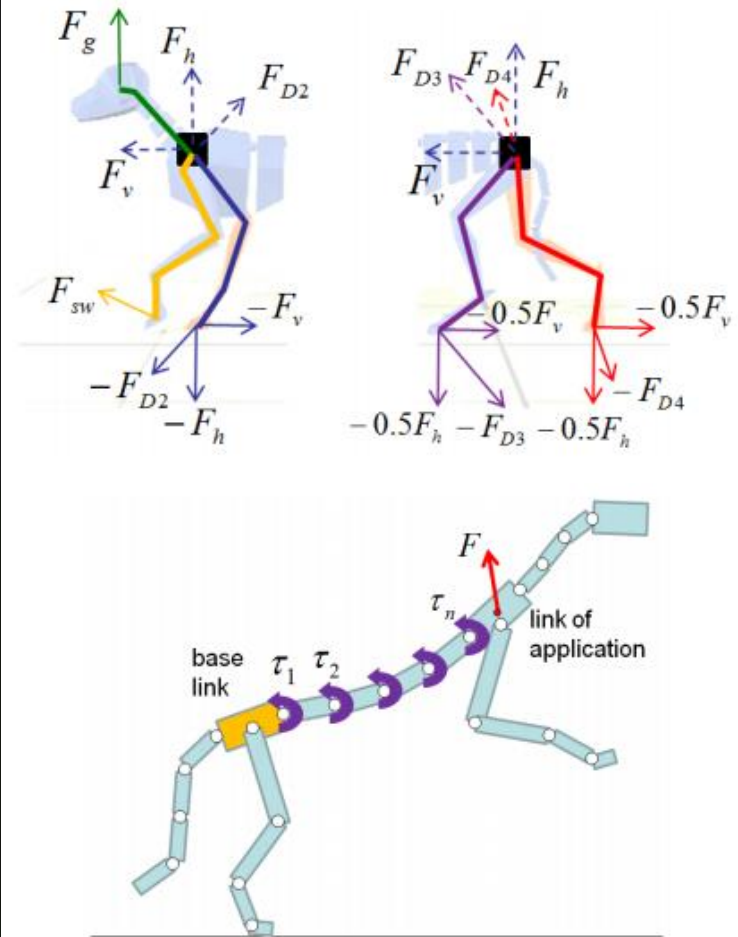


[MIT Biomimetic Robotics Lab 2019]

# Manual Controller Design

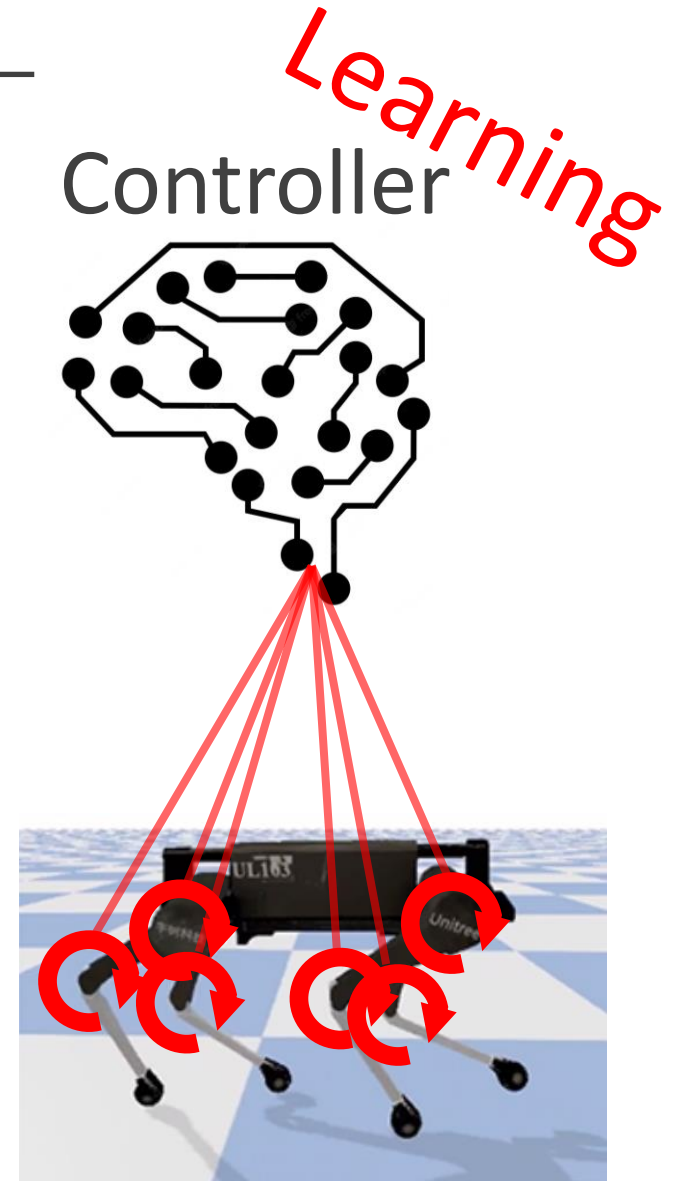
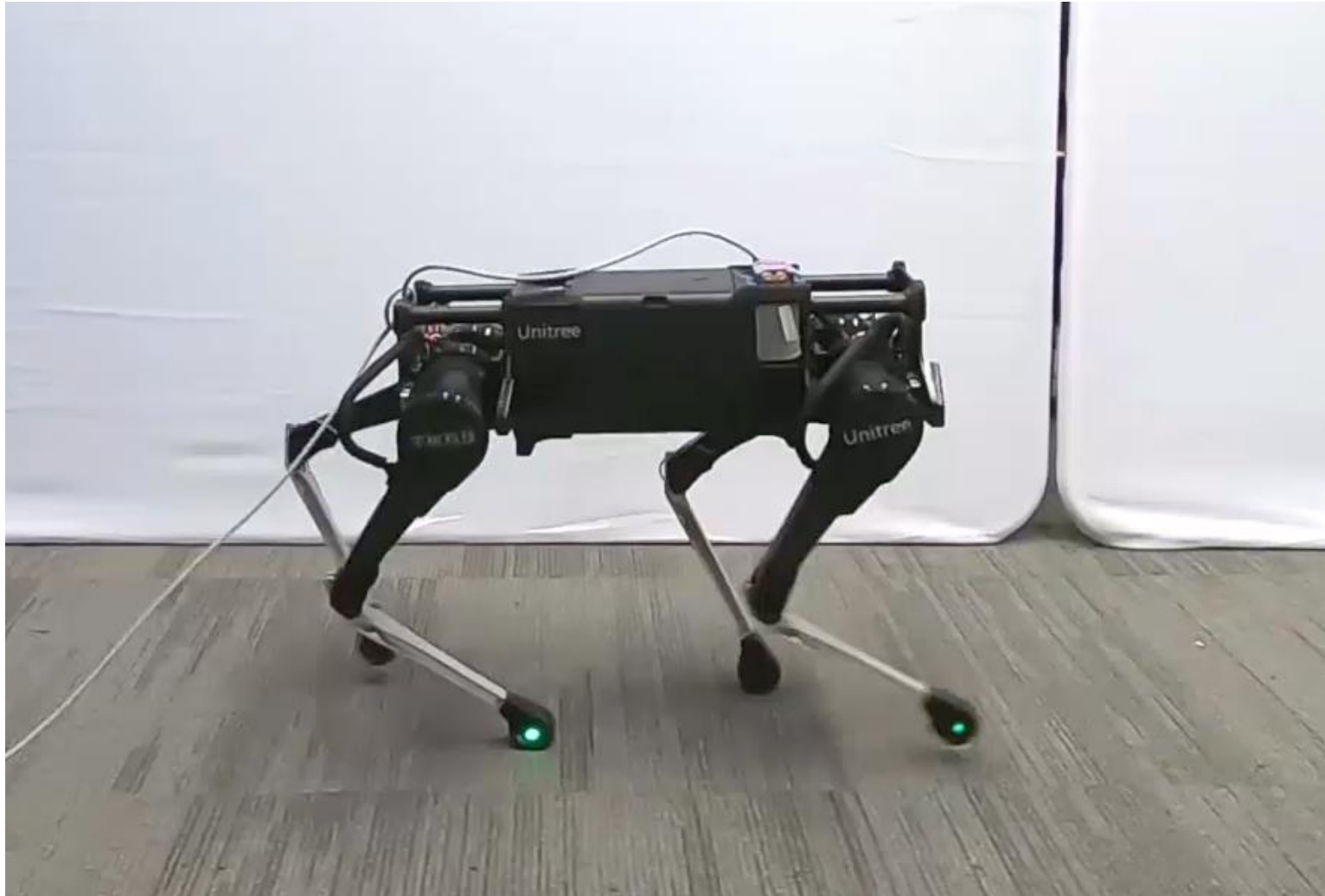


[Coros et al., 2011]



# Decision Making Problems

---

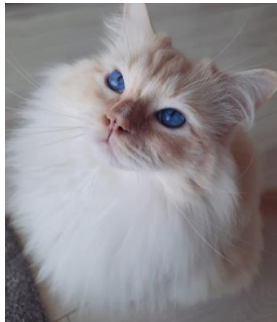




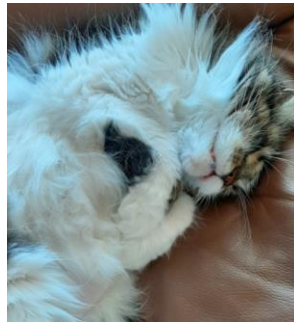
# ML Paradigms

## Supervised Learning

$$\{(\mathbf{x}_i, y_i)\}$$



Cat



Cat



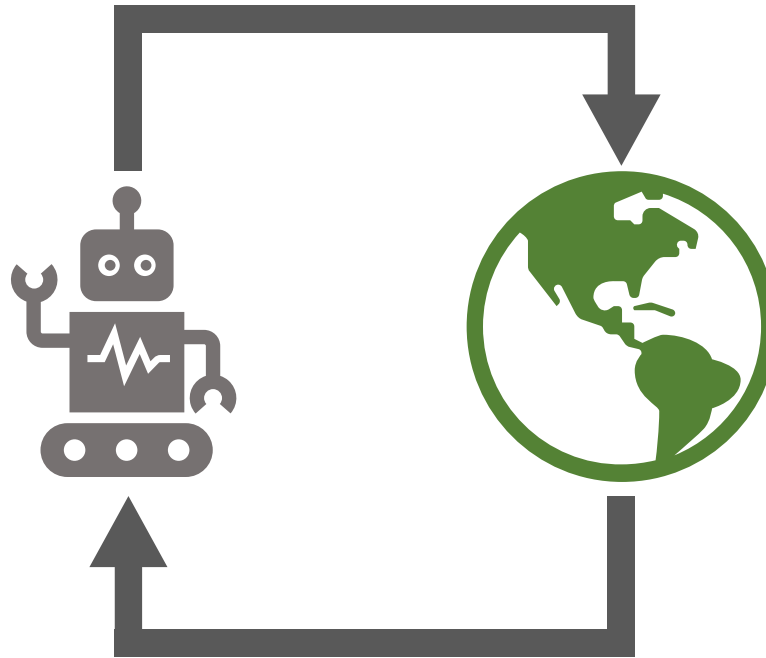
Dog



Dog

## Reinforcement Learning

$$\{(\mathbf{x}_i, y_i, r_i)\}$$



## Unsupervised Learning

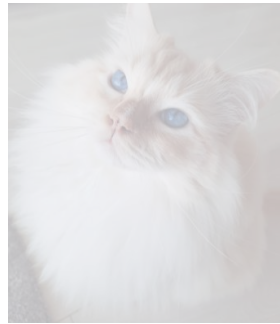
$$\{\mathbf{x}_i\}$$



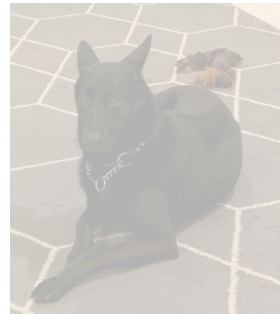
# ML Paradigms

**Supervised Learning**

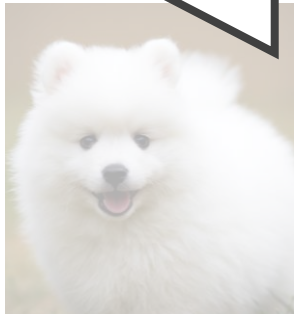
$$\{(\mathbf{x}_i, y_i)\}$$



Cat



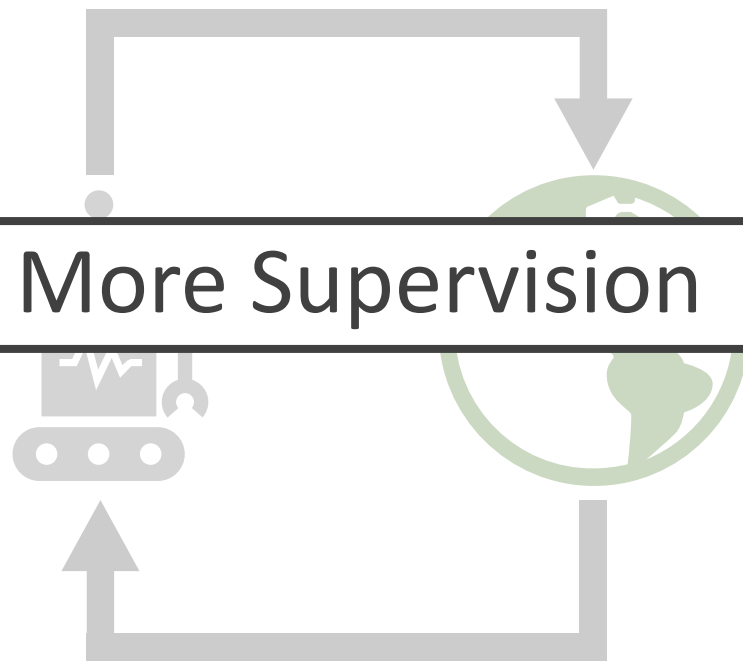
Dog



Dog

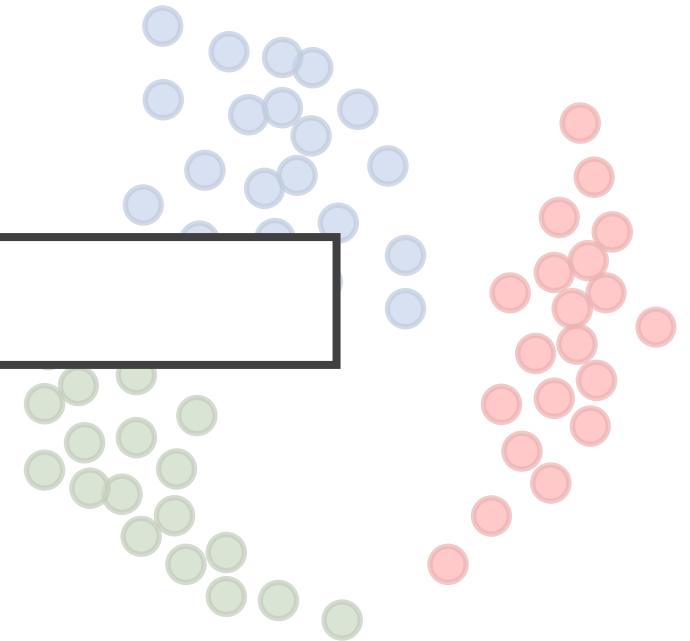
**Reinforcement Learning**

$$\{(\mathbf{x}_i, y_i, r_i)\}$$



**Unsupervised Learning**

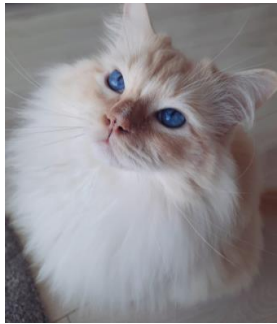
$$\{\mathbf{x}_i\}$$



# ML Paradigms

## Supervised Learning

$$\{(\mathbf{x}_i, y_i)\}$$



Cat



Cat



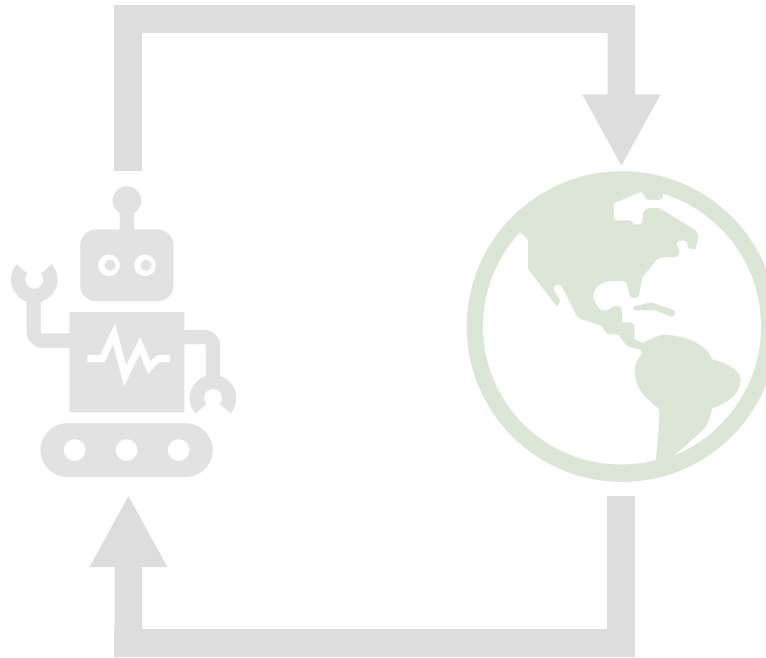
Dog



Dog

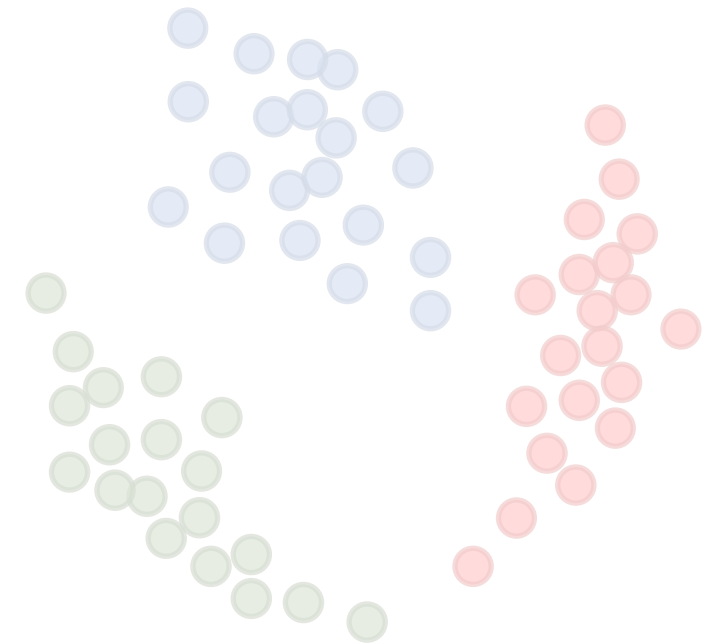
## Reinforcement Learning

$$\{(\mathbf{x}_i, y_i, r_i)\}$$



## Unsupervised Learning

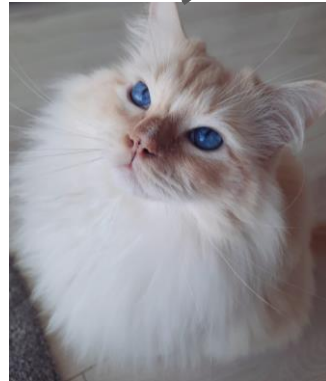
$$\{\mathbf{x}_i\}$$



# Supervised Learning

---

$\{(\underline{x}_i, \underline{y}_i)\}$



“Cat”

# Supervised Learning

---

$\{(\underline{x}_i, \underline{y}_i)\}$

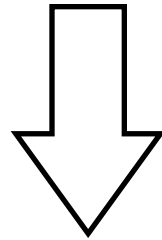


“Dog”

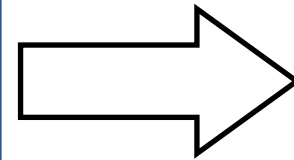
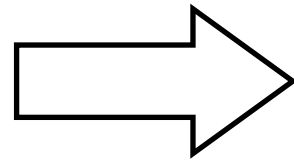
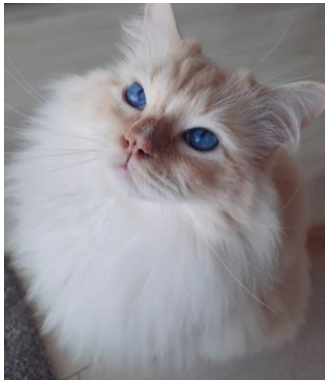
# Supervised Learning

---

$$\{(\mathbf{x}_i, y_i)\}$$



$$f(y_i | \mathbf{x}_i)$$



“Cat”

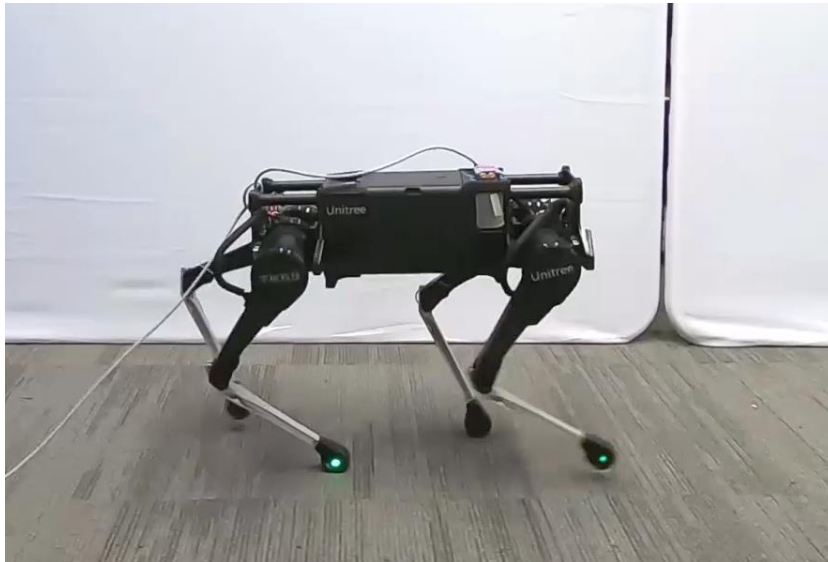
# Supervised Learning

---

$$\{(\underline{x}_i, \underline{y}_i)\}$$

Robot State

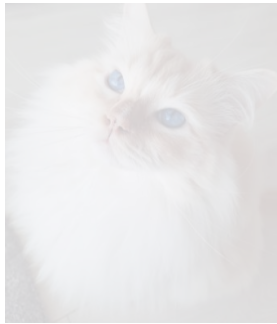
Motor Commands



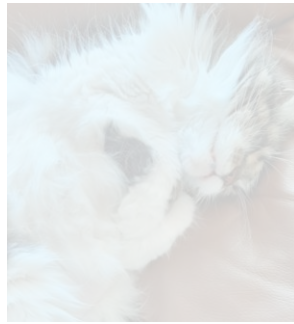
# ML Paradigms

## Supervised Learning

$$\{(\mathbf{x}_i, y_i)\}$$



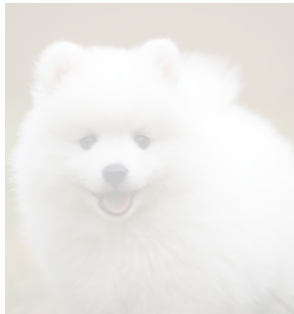
Cat



Cat



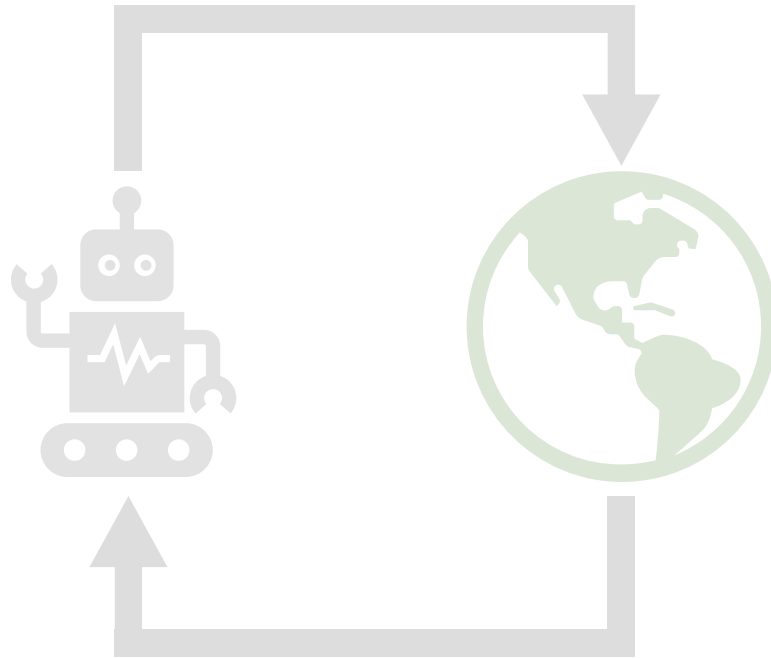
Dog



Dog

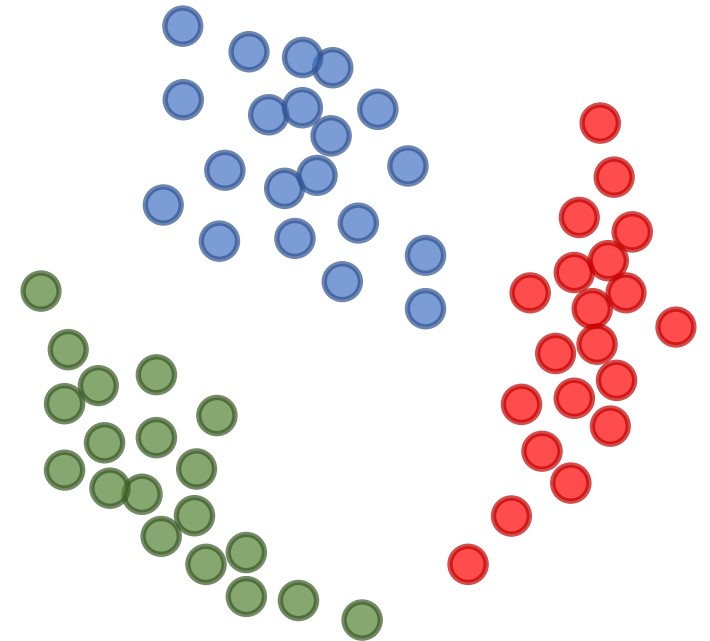
## Reinforcement Learning

$$\{(\mathbf{x}_i, y_i, r_i)\}$$



## Unsupervised Learning

$$\{\mathbf{x}_i\}$$

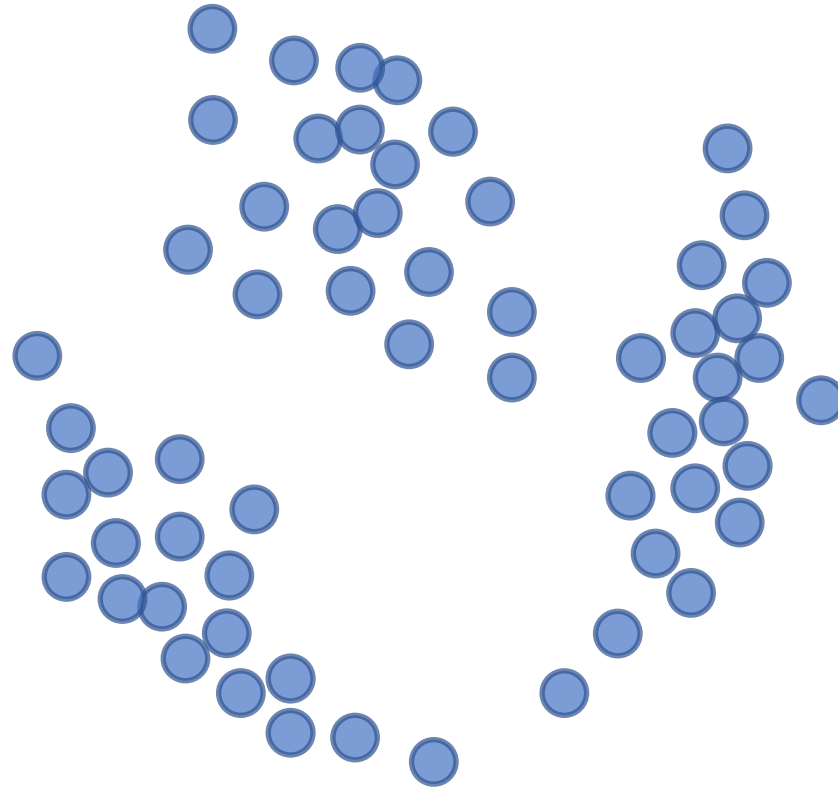




# Unsupervised Learning

---

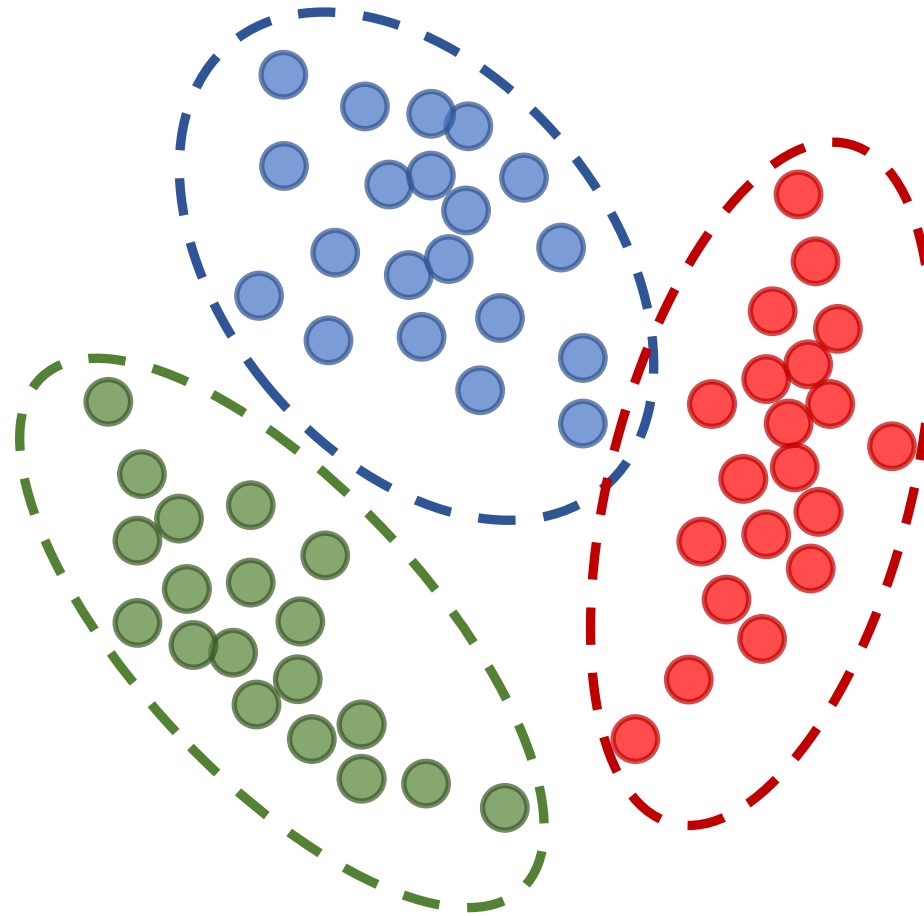
$$\{\mathbf{x}_i\}$$



# Unsupervised Learning

---

$\{\mathbf{x}_i\}$



# Unsupervised Learning

---

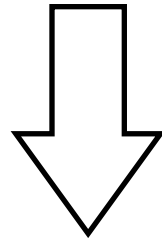
$\{\mathbf{x}_i\}$



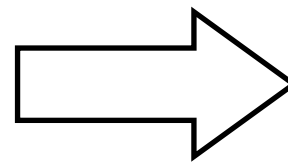
# Unsupervised Learning

---

$\{\mathbf{x}_i\}$



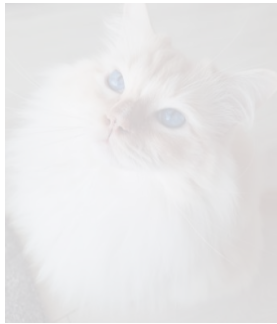
$$f(\mathbf{x}) \approx p(\mathbf{x})$$



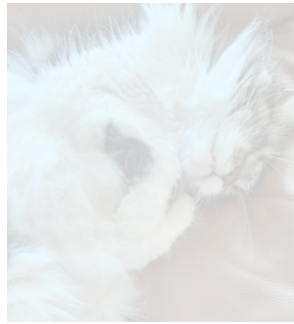
# ML Paradigms

## Supervised Learning

$$\{(\mathbf{x}_i, y_i)\}$$



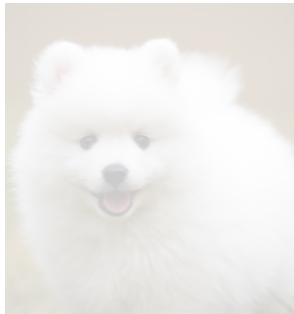
Cat



Cat



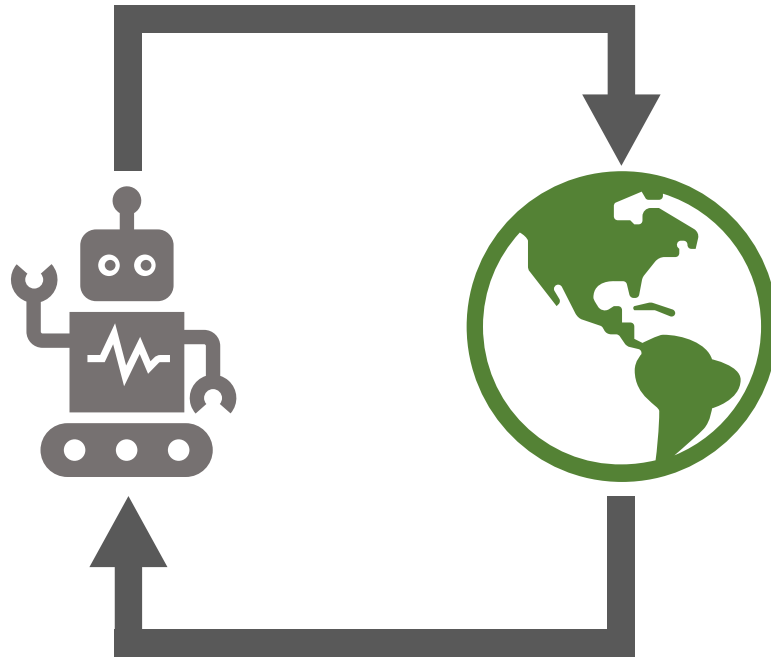
Dog



Dog

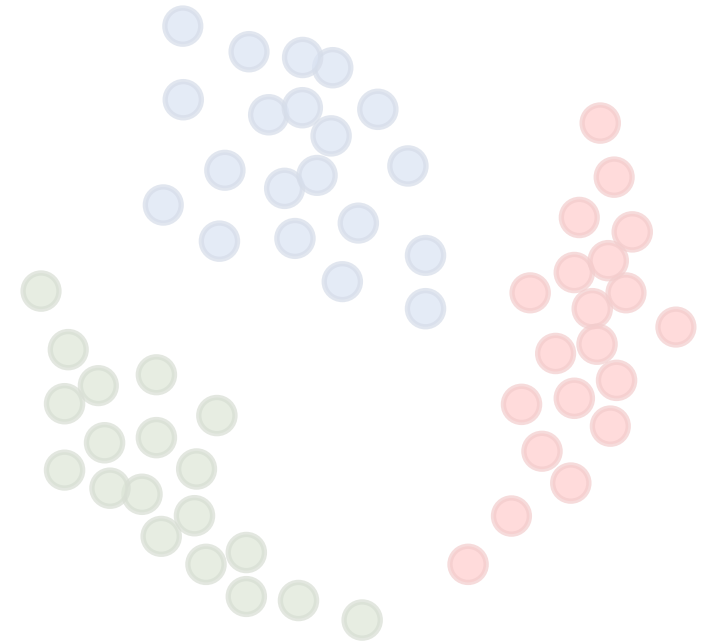
## Reinforcement Learning

$$\{(\mathbf{x}_i, y_i, r_i)\}$$



## Unsupervised Learning

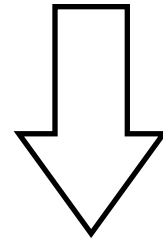
$$\{\mathbf{x}_i\}$$



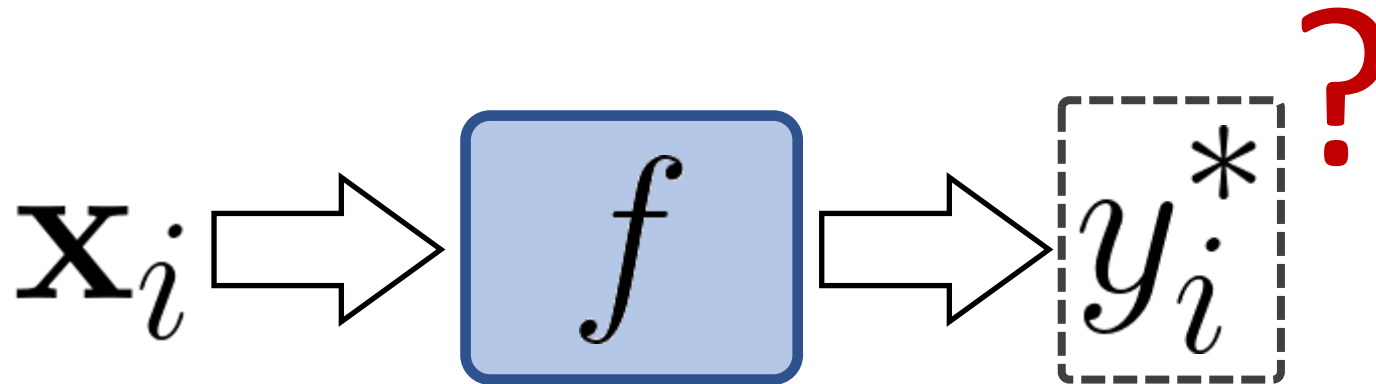
# Reinforcement Learning

---

$$\{(\mathbf{x}_i, y_i, r_i)\}$$



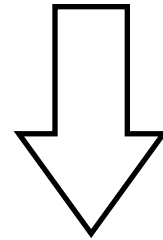
$$f(y_i | \mathbf{x}_i)$$



# Reinforcement Learning

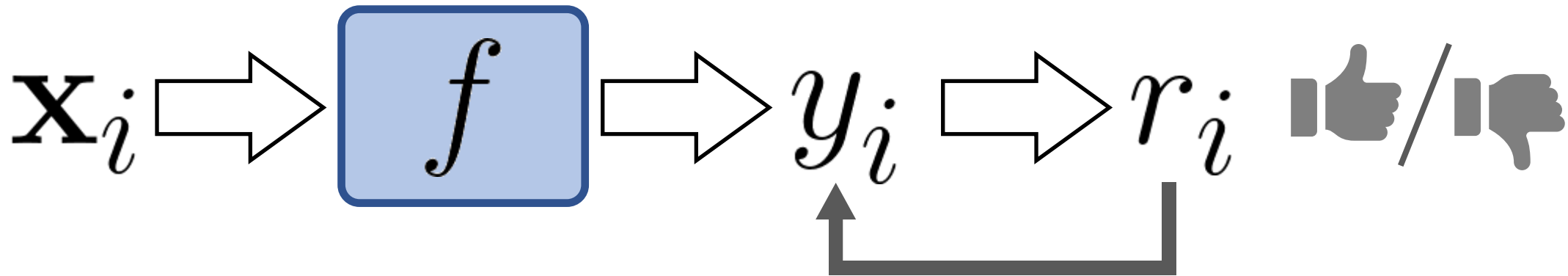
---

$$\{(\mathbf{x}_i, y_i, \underline{r_i})\}$$



Score/Reward

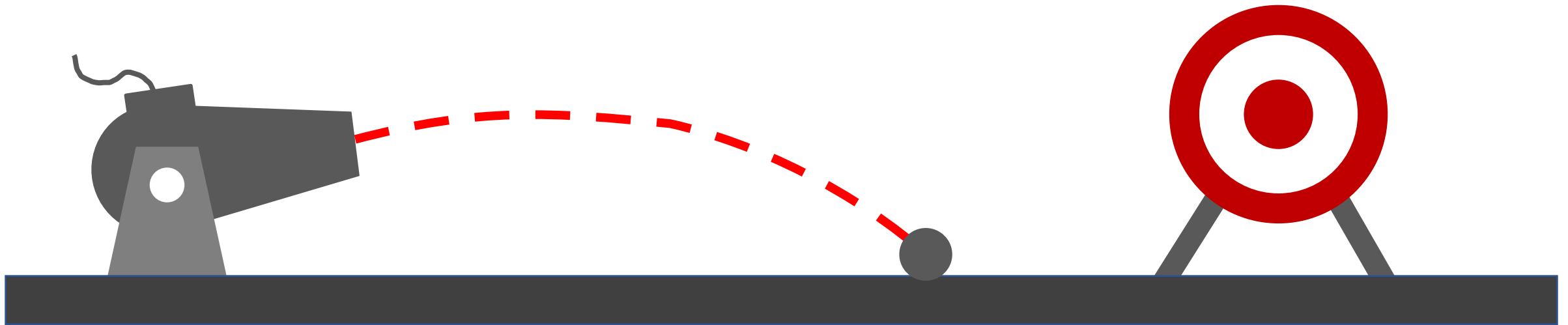
$$f(y_i | \mathbf{x}_i)$$



# Reinforcement Learning

---

- Learning through trial-and-error





# Reinforcement Learning

---

- Learning through trial-and-error



# Reinforcement Learning

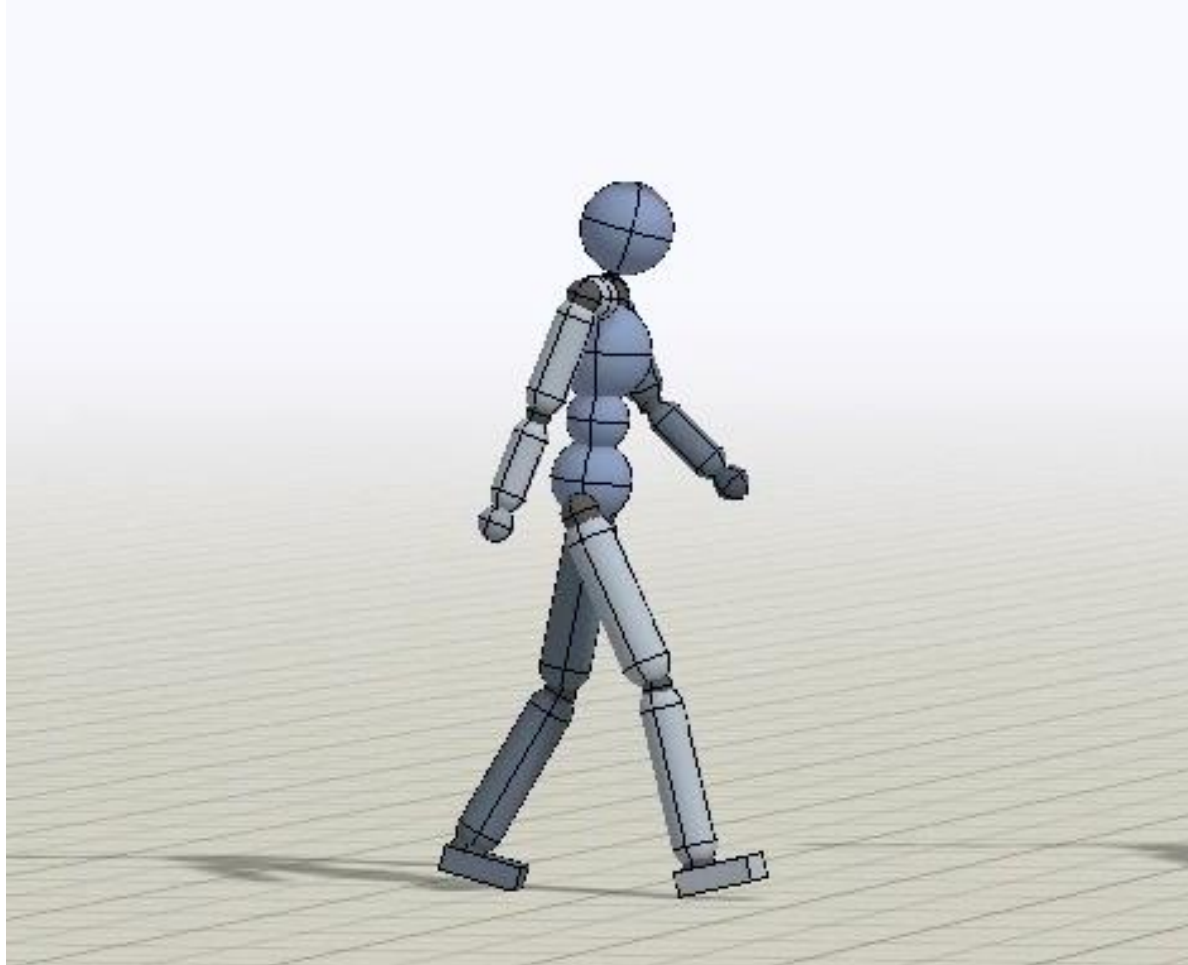
---

- Learning through trial-and-error



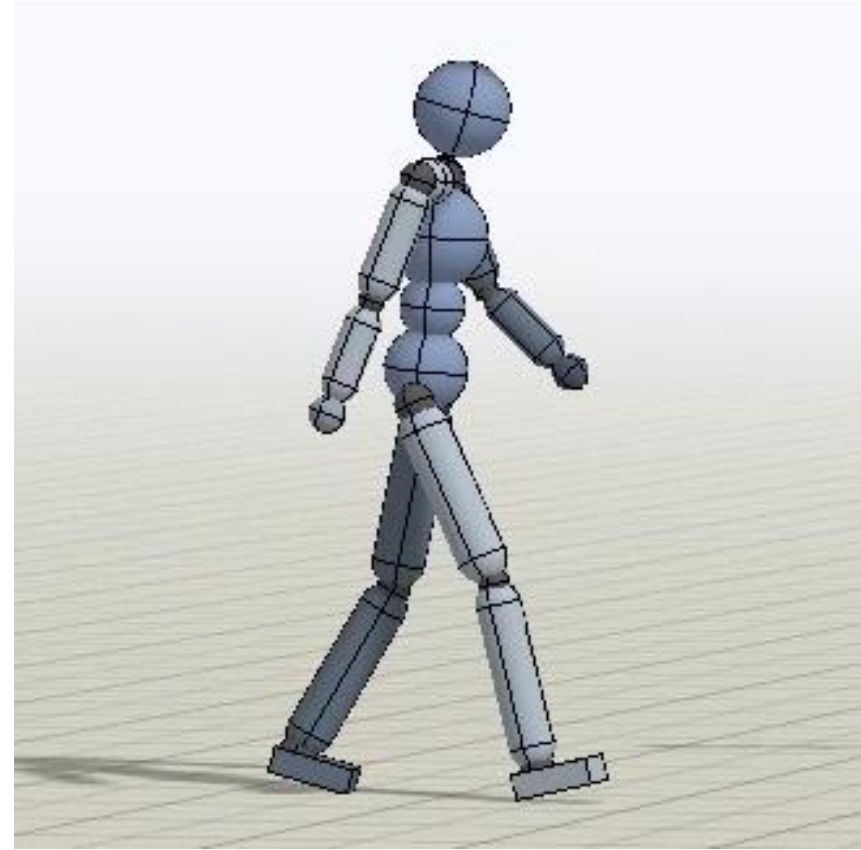
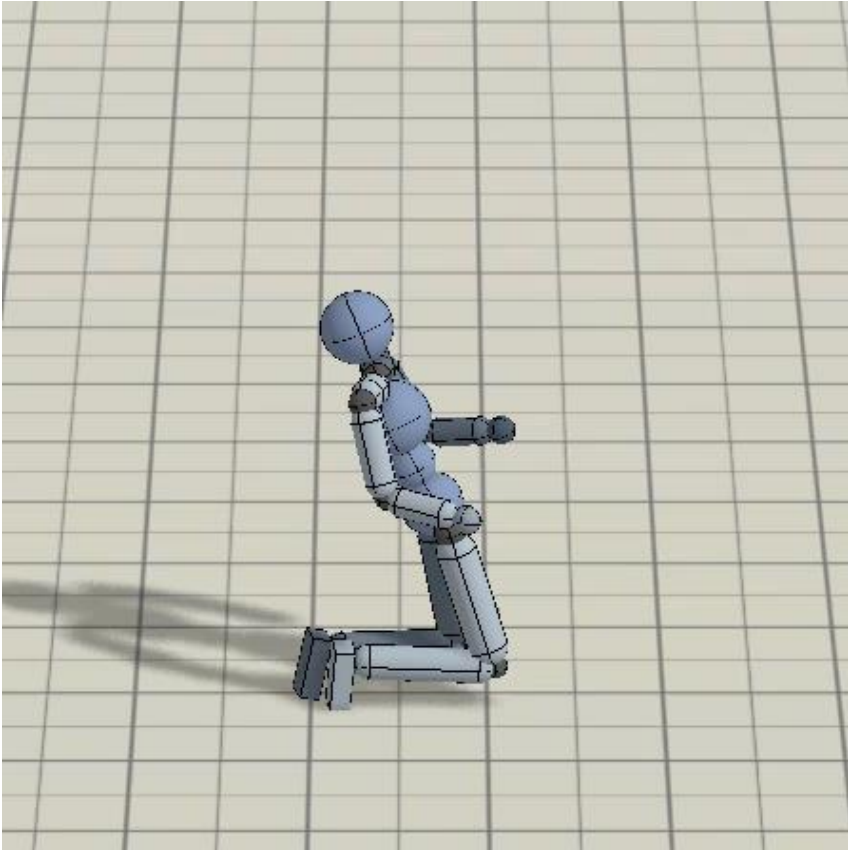
# Reinforcement Learning

---



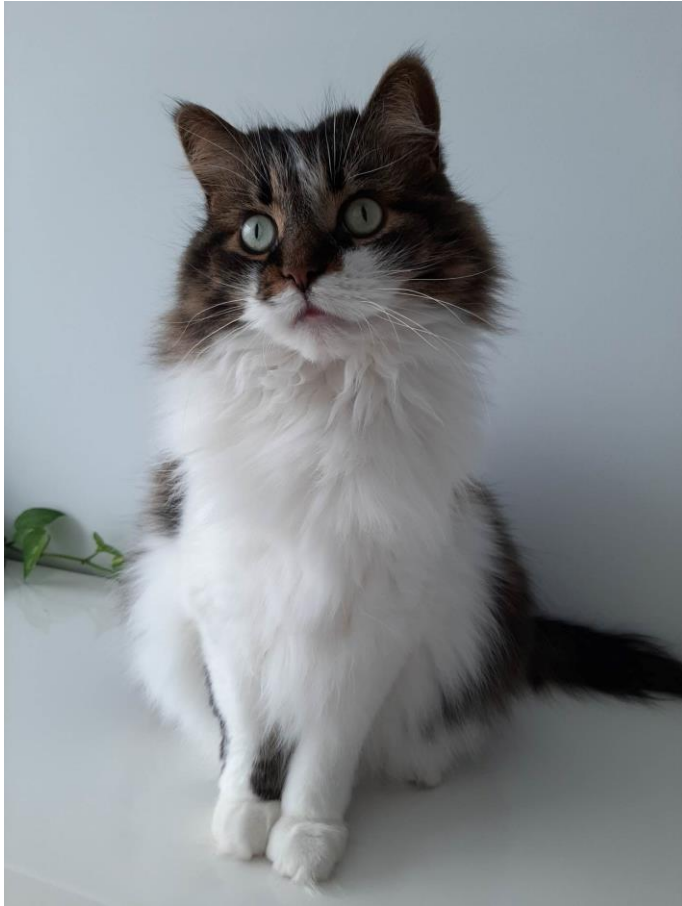
# Reinforcement Learning

---

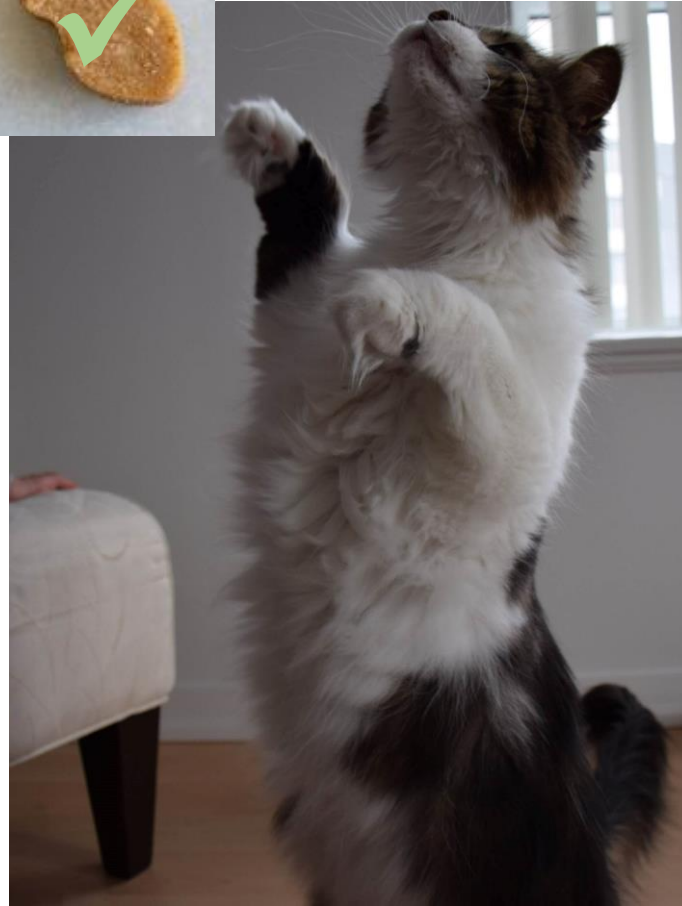


# Reinforcement Learning

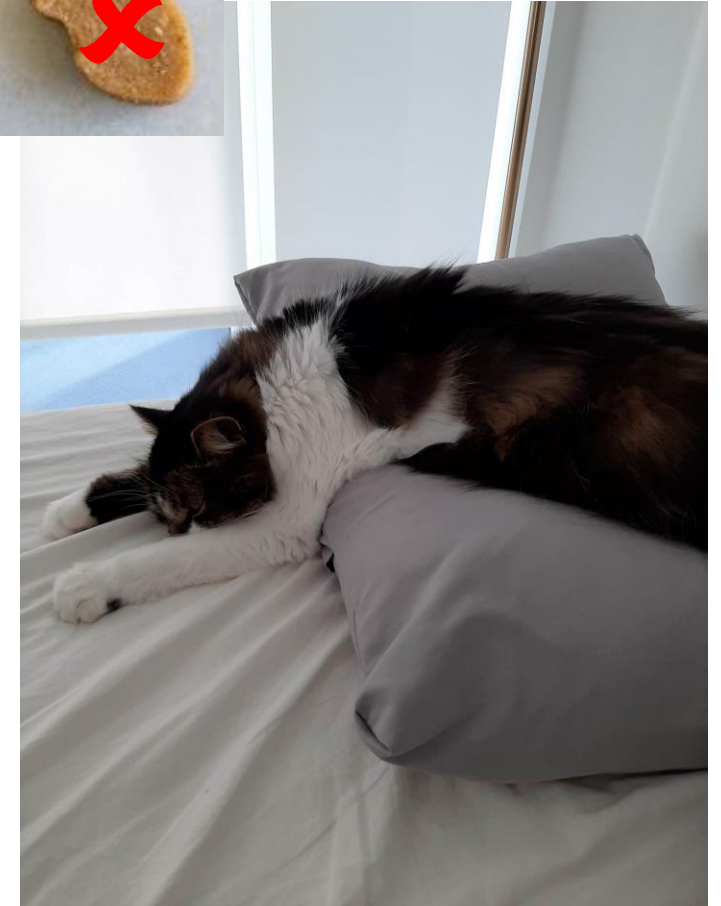
---



Mellow



Reward



Punishment

# Reinforcement Learning

---

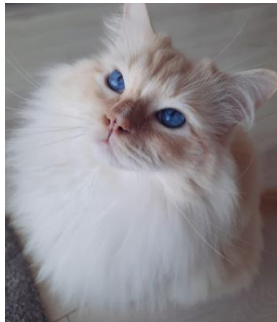


[AlphaGo 2016]

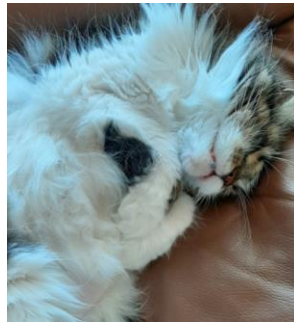
# Data Sources

## Supervised Learning

$$\{(\mathbf{x}_i, y_i)\}$$



Cat



Cat



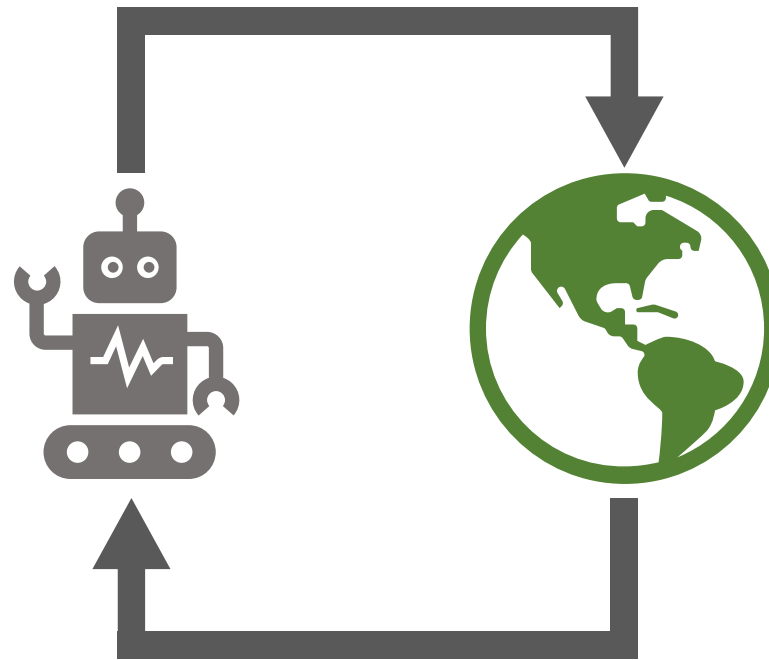
Dog



Dog

## Reinforcement Learning

$$\{(\mathbf{x}_i, y_i, r_i)\}$$



## Unsupervised Learning

$$\{\mathbf{x}_i\}$$

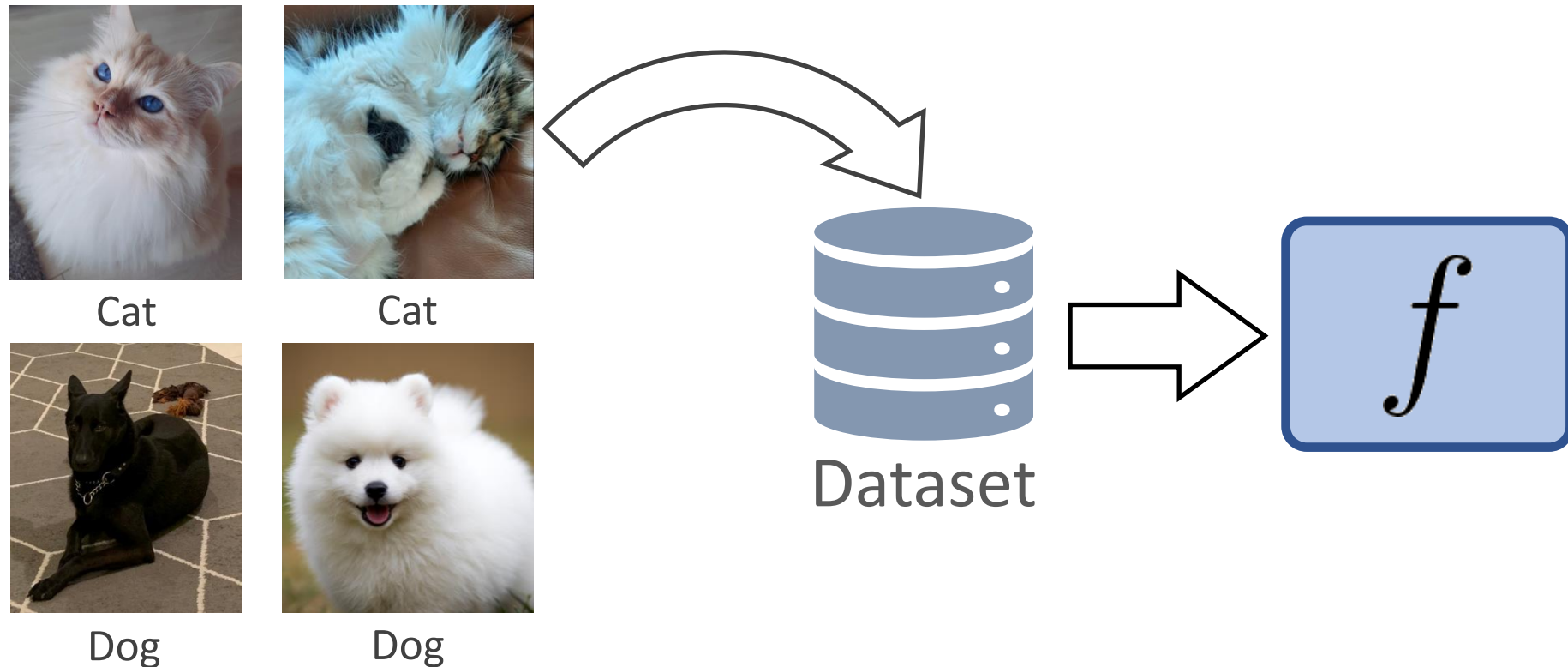


# Passive Learning

---

**Passive Learning:** Agent is given a fixed dataset to learn from

- Agent passively observes the world
- does not affect its environment





# Active Learning

---

**Active Learning:** Agent collects its own data

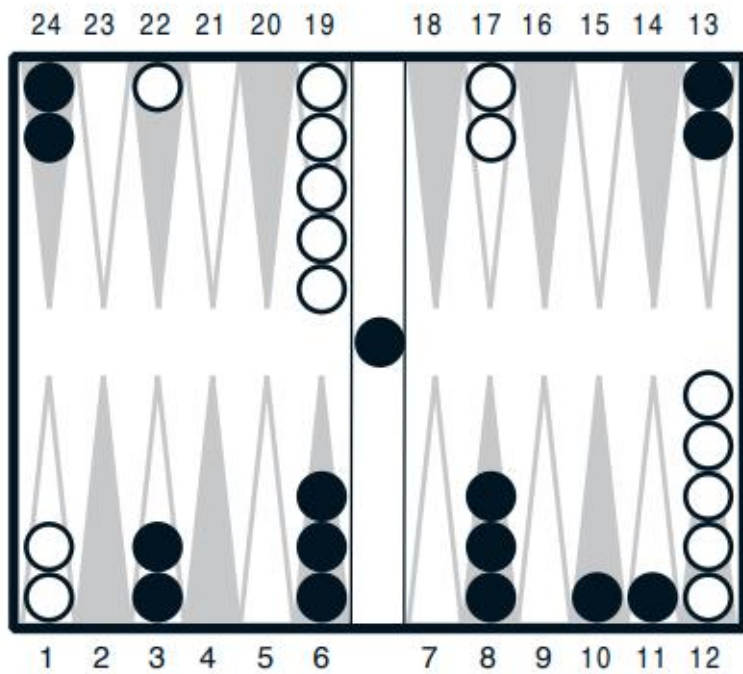
- Agent interact and affects its environment
- Data depends on the agent's behaviors



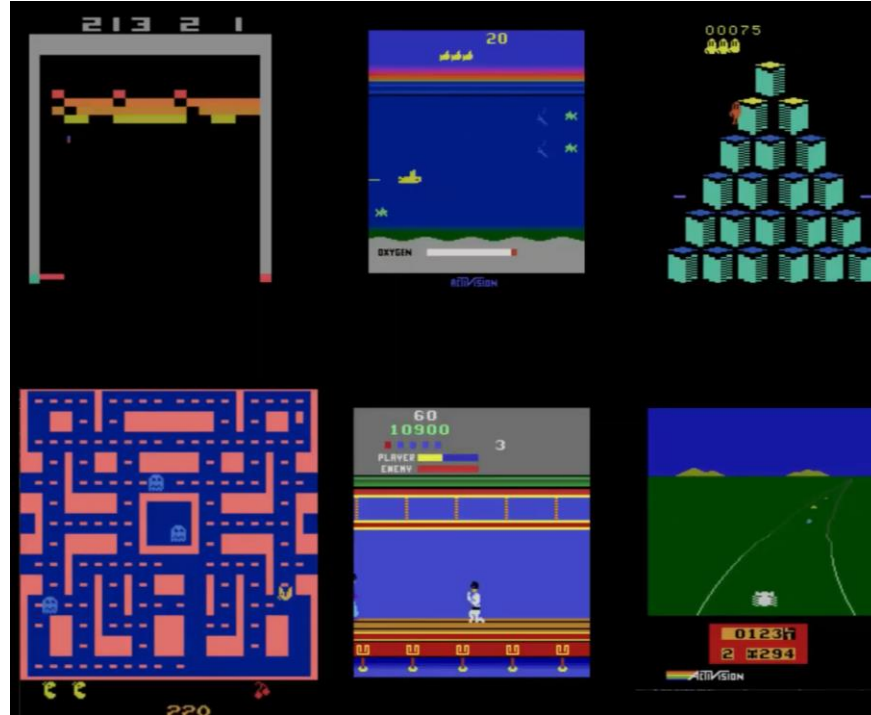
# Applications

# Games

---



[Tesauro 1995]



[Mnih et al. 2015]

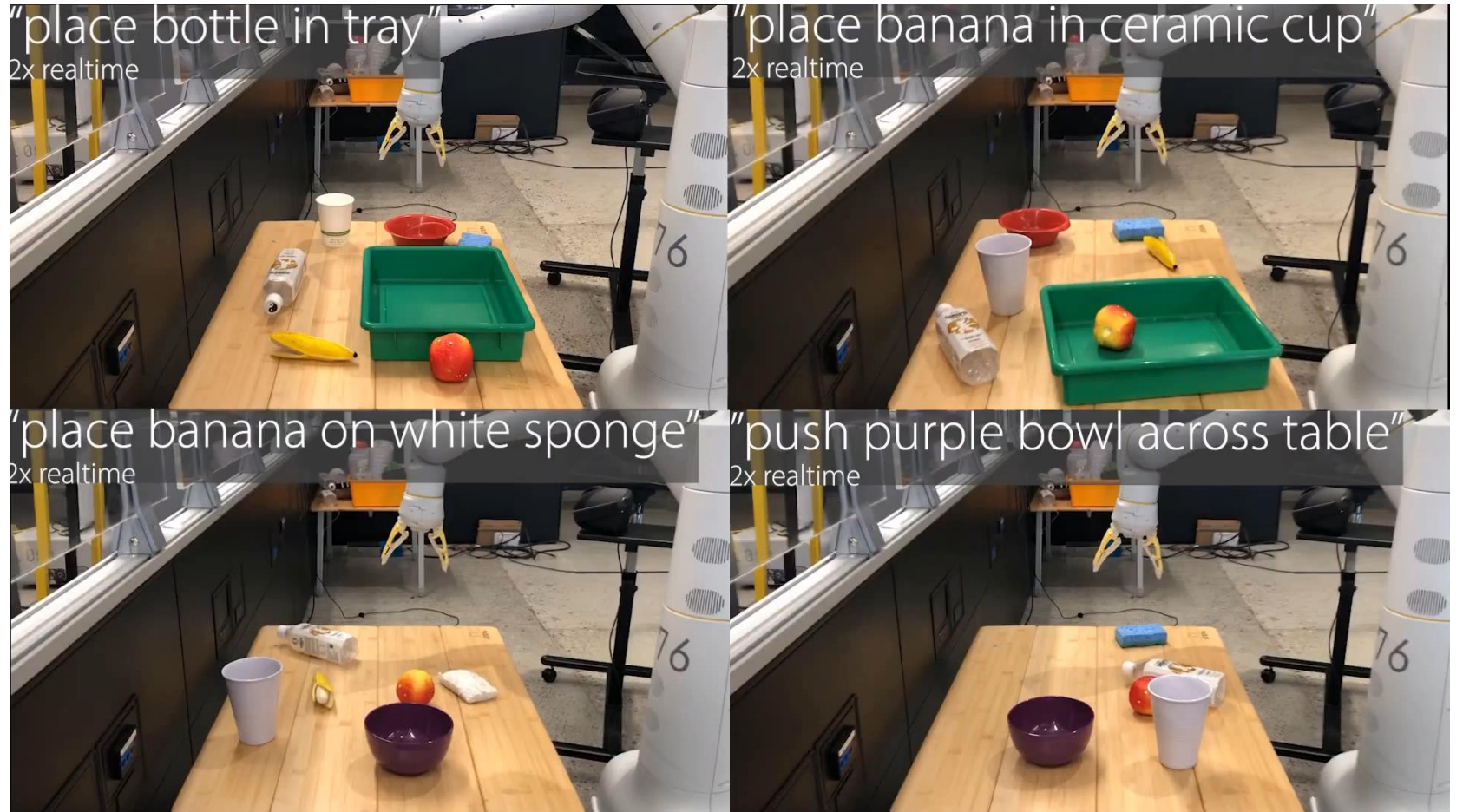
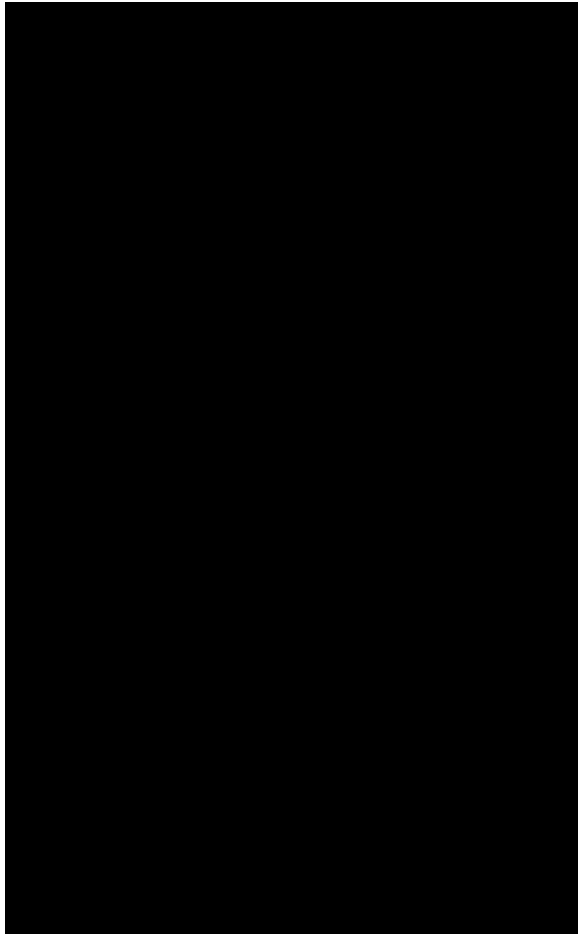


[Silver 2017]

Grandmaster Level in StarCraft II Using Multi-Agent Reinforcement Learning

[Vinyals 2019]

# Robotic Manipulation



[Nagabandi et al. 2019]

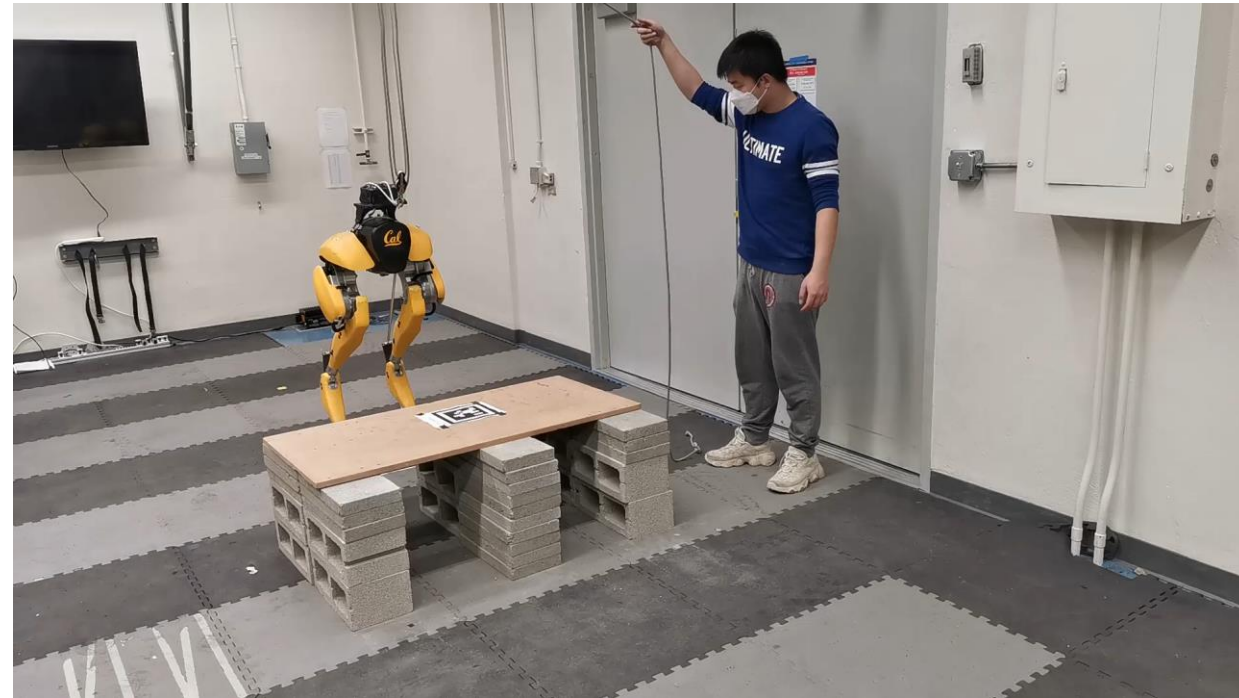
[Jang et al. 2021]

# Robotic Locomotion

---



[Miki et al. 2022]



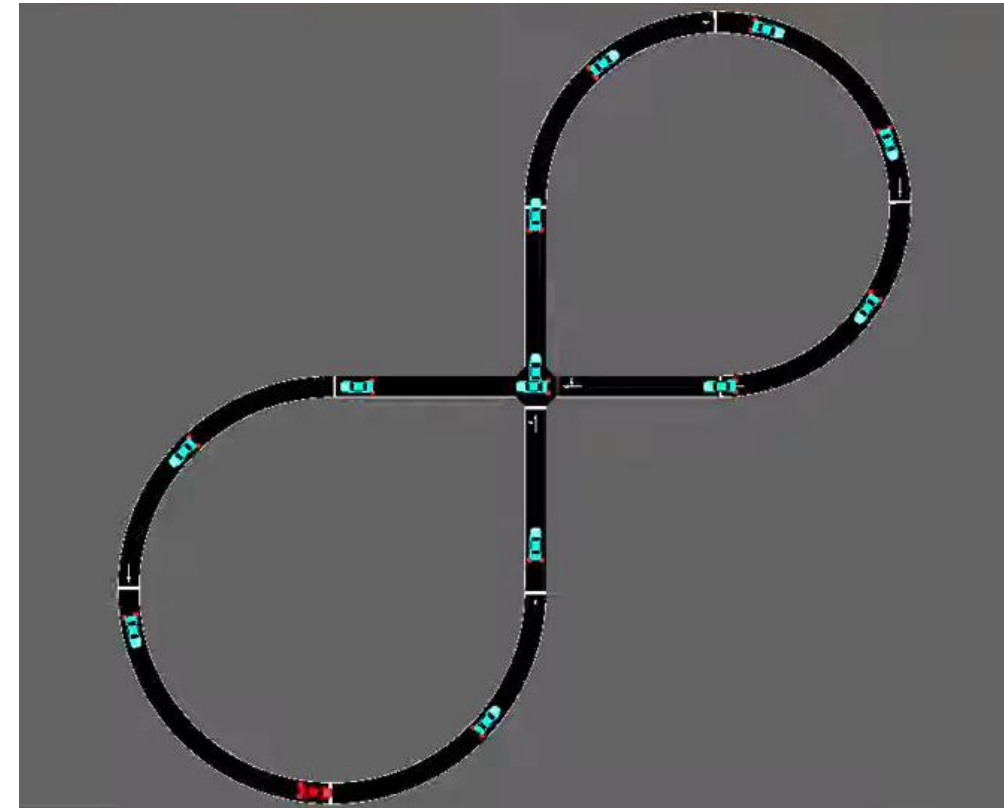
[Li et al. 2023]

# Autonomous Driving

---



[Bojarski et al. 2016]



[Wu et al. 2021]

# Energy Conservation




Safety-First AI for Autonomous Data Centre Cooling and Industrial Control  
[Gamble and Gao 2018]



# Recommendation Systems

## Customers who viewed this item also viewed

Page 1 of 6



UDDHAV GOLD Pure Copper Bottle for Water 1 Liter Dirt Proof Leak Proof and Joint Less Ayurveda...  
★★★★★ 53  
₹ 530.00

Ayurveda Copper™ |Copper Modern Art Printed and Matt Finish Antique Yoga Water...  
★★★★☆ 87  
₹ 499.00 ✓prime

UDDHAV GOLD Pure Copper Bottle for Water 1 Liter Mat Finish Dirt Proof Leak Proof and Joint...  
★★★★★ 53  
₹ 525.00

UDDHAV GOLD Pure Copper Bottle for Water 1 Liter Dirt Proof Leak Proof and Joint Less Ayurveda...  
★★★★★ 53  
₹ 619.00 ✓prime

Ayurveda Copper™ Pure Copper Bottle (Meena Black Gold Spiral Artwork, 1000 ml)(Pack of 2) for...  
★★★★☆ 87  
₹ 999.00 ✓prime

Ayurveda Copper™ |Copper's Pure Copper Printed Water Bottle | Designer Copper Bottle...  
★★★★☆ 87  
₹ 499.00 ✓prime

Just Copper Combo Pack of Pure Copper Modern Art Printed and Outside Lacquer Coated Bottle,...  
★★★★☆ 349  
₹ 410.00

## Sponsored products related to this item

Page 1 of 158



Indian Art Villa Steel Copper Jug Pitcher with 2

Amazon Brand - Solimo Copper Hammered Jug

LCLLOTUS Copper Water Bottle Joint Free and Leak

Femora Borosilicate Glass Tea Pot Carafe with

JaipurCrafts Pure Copper Modern Art Printed and

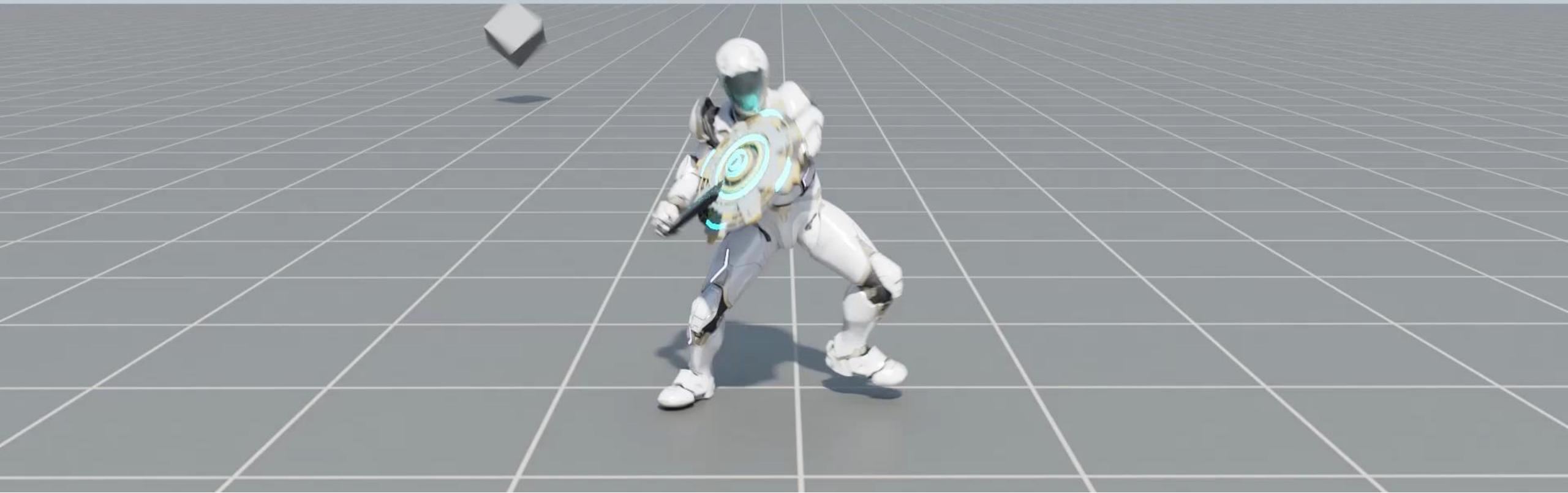
Machak Spiral Glass Water Jug with Lid Beverage

COPPERTOWN Handmade Lacquer Coated 950 ml

Reinforcement Learning to Optimize Long-term User Engagement in Recommender Systems  
[Zou et al. 2019]

# Computer Graphics

---



ASE: Large-Scale Reusable Adversarial Skill Embeddings for Physically Simulated Characters  
[Peng et al. 2022]

# Logistics

# Preliminaries

---

- There will be **a lot** of math
  - Probability theory
  - Calculus
  - Linear algebra
- Machine learning
  - Neural networks
  - Optimization
  - Supervised learning
  - Unsupervised learning
- Programming
  - Python
  - PyTorch

# Lectures

---

**00:** Introduction

**01:** MDP

**02:** Policy Evaluation

**03:** Behavioral Cloning

**04:** Policy Search

**05:** RL Algorithms

**06:** Policy Gradient

**07:** Q-Learning

**08:** Actor-Critic Algorithms

**09:** Model-Based RL

**10:** On-Policy vs. Off-Policy Algorithms

**11:** Advance Policy Gradient

**12:** Advance Q-learning

**13:** Exploration

**14:** Unsupervised RL

**15:** Imitation Learning

**16:** Domain Transfer

**17:** Offline RL

\*Tentative

# Grading

---

- 3 programming assignments (10% each)
- Paper presentation (20%)
- Course project (50%)
  - Proposal (10%)
  - Presentation (20%)
  - Report (20%)
- No exams

# Paper Presentation

---

- Present an RL-related paper
- Groups 3-4 (depending on class size)

# Course Project

---

- Apply reinforcement learning to solve an interesting problem
  - No board games
  - No Atari games
  - No standard benchmark problems (OpenAI gym, DeepMind Control Suite)
- Groups 3-4 (depending on class size)
- 1-2 page proposal due in mid semester
- Project presentations at the end of the semester
- Project report due at the end of the semester



# Course Page

---

## CMPT 729: Reinforcement Learning



Reinforcement learning is the branch of machine learning that studies learning to act. Agents observe, predict, and act to change their environment. Reinforcement learning has notable success in learning to play games and control robots. In this course, we will cover fundamental concepts and algorithms, and introduce techniques that underlie many of the successes from reinforcement learning.

**Instructor:** Jason Peng

**TA:** Sha Hu

**Lectures:**

**Wed** 11:30am-12:20pm (SWH10051)

**Fri** 10:30am-12:20pm (AQ5016)

---

## Grading

**3 programming assignments (30%)**

- A1 (10%) - Due Jan 28
- A2 (10%) - Due Feb 25
- A3 (10%) - Due Mar 10

[[xbpeng.github.io/teaching/cmpt\\_729/](http://xbpeng.github.io/teaching/cmpt_729/)]

# Discussion Forum

## CMPT 729 G1: Discussion Forum

[\[Forum Summary\]](#)

[\[New Thread\]](#)

*Nothing posted yet.*

[\[Discussion forum identities\]](#)

[\[Activity digest\]](#)

### Unanswered Questions

*None*

A question is considered “answered” if (1) an instructor/TA has replied, (2) an instructor/TA has reacted positively (👍, ❤️, 🍌) to a student reply, or (3) the question-asker has marked it answered or reacted positively to a reply.

### Unread Activity

*None*

### Search Posts

Search posts:

# Office Hours

---

**Jason:** Wednesday 3-4pm in TASC 9213

**Sha:** Friday 3-4pm in ASB 9808

# Summary

---

- What is reinforcement learning?
- Applications
- Logistics