JOHNS HOPKINS
CAREY BUSINESS SCHOOL

# Lecture 7

## BU.330.775 Machine Learning

Minghong Xu, PhD.
Associate Professor

# Review

- Clustering vs classification
- Use cases
  - Customer segmentation
  - Anomaly detection
  - Image segmentation
  - Image search engine
- K-means clustering: iterative algorithm

# Today’s Agenda

- Reinforcement Learning
- Final Review

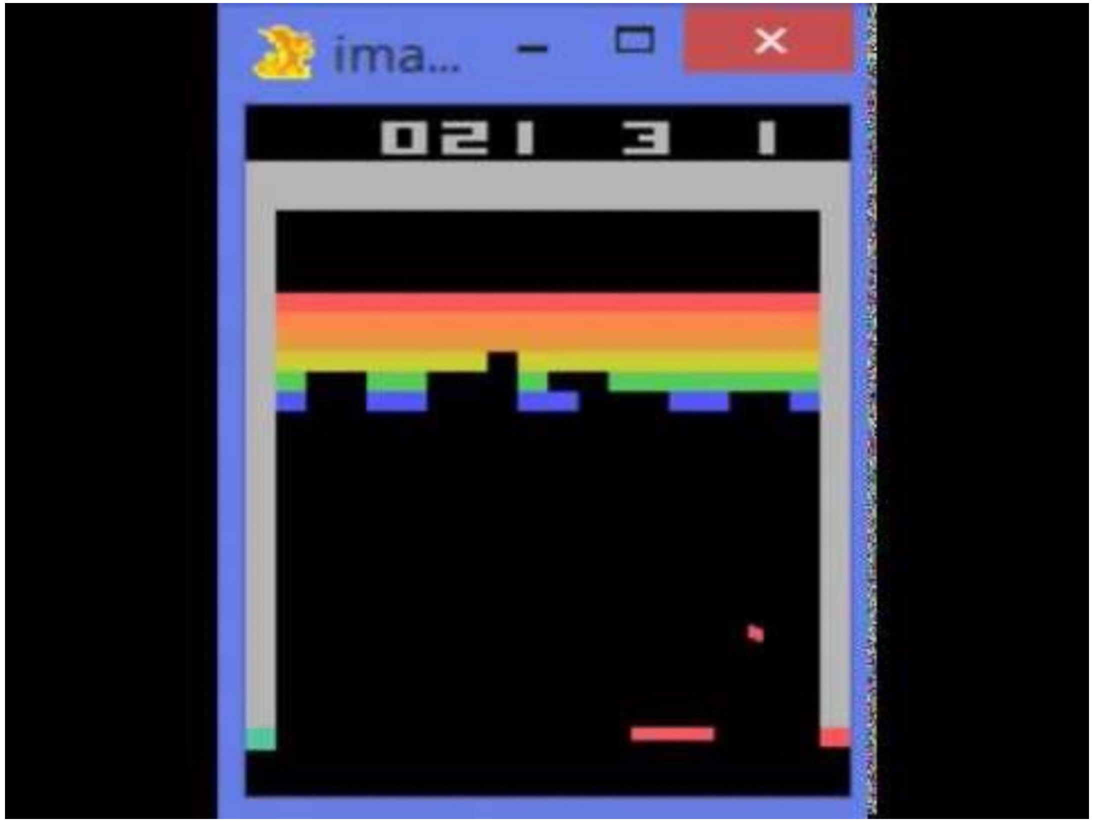
# History of Reinforcement Learning

- Born 1950s
- Concept of trial-and-error learning: learn from failures
- Bellman equation
  - Value of a state is equal to the immediate reward obtained in that state, plus the expected value of the next state
  - Recursive way of solving a decision problem

Notes

# Breakthrough

- DeepMind Atari game (2013)

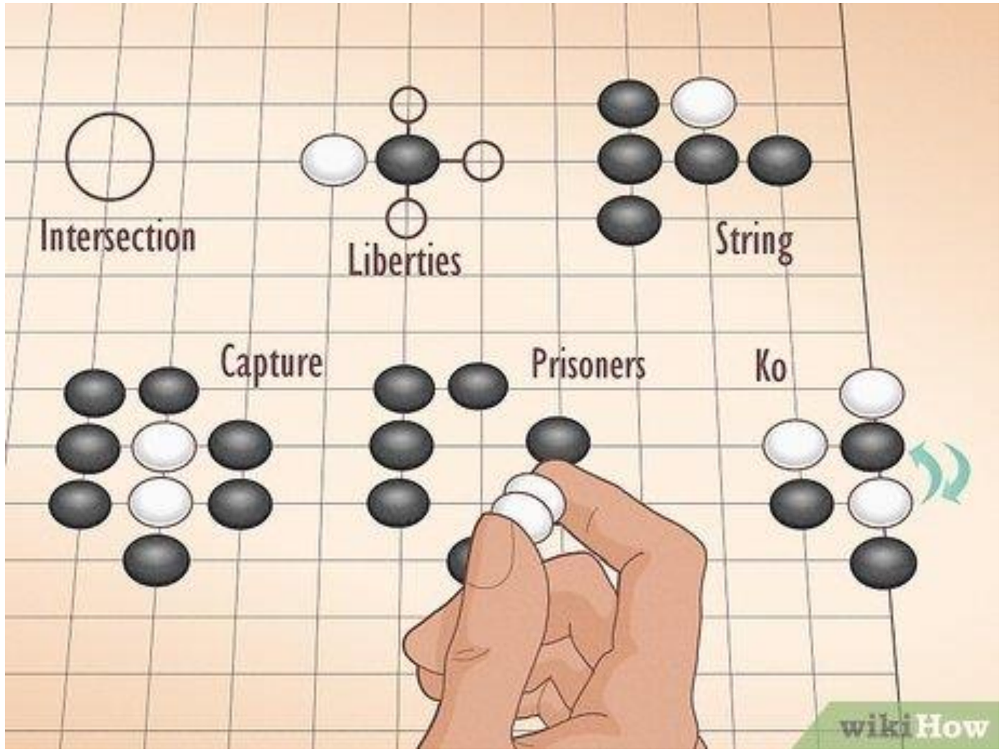


# DeepMind AlphaGo (2016)

- Apply power of deep learning to reinforcement learning

# Reinforcement Learning

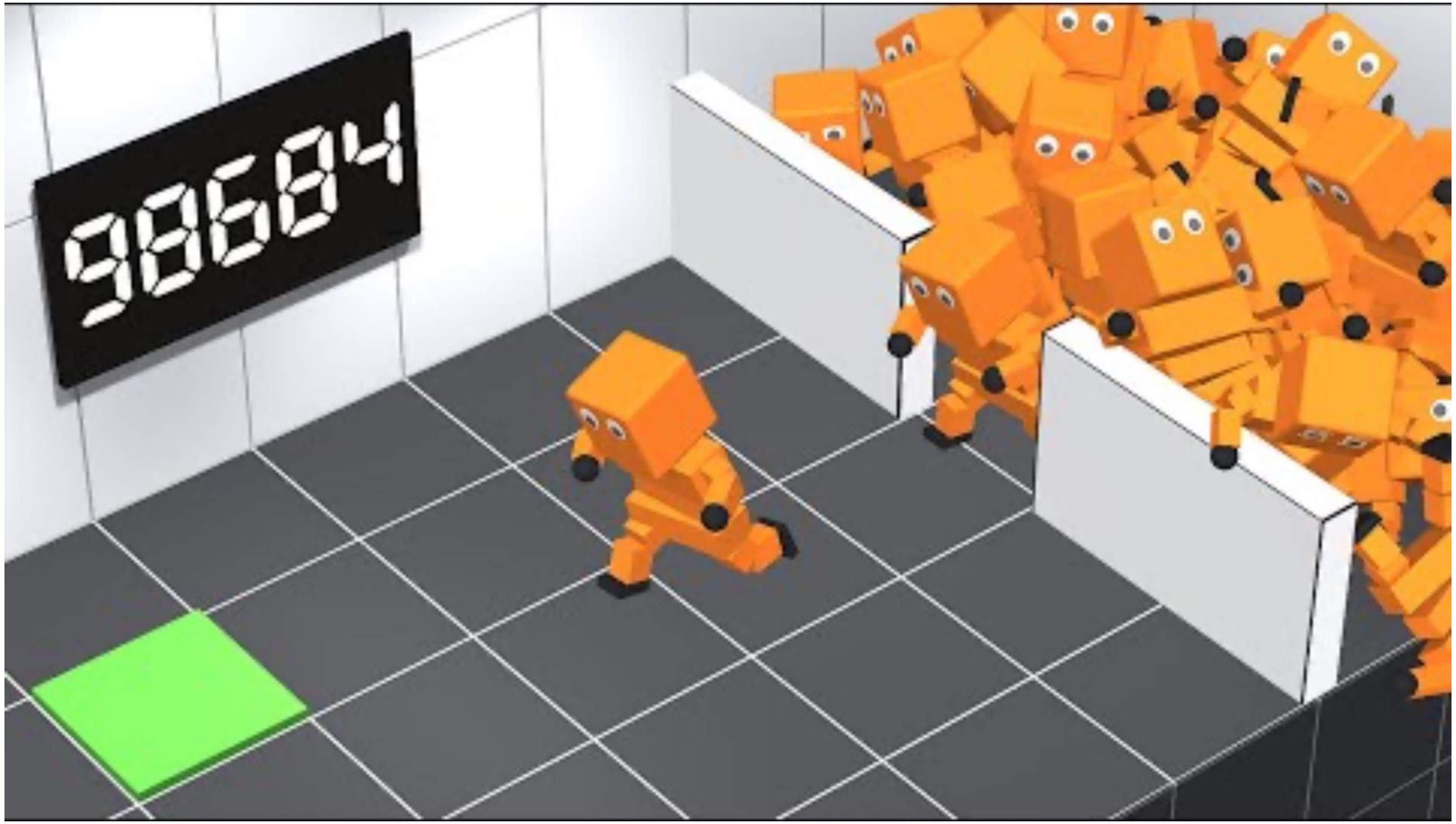
- A software **agent** makes observations and take **actions** within an environment, and in return it receives **rewards** from the environment
- Objective: **maximize** its expected rewards over time



# Policy

- Algorithm a software agent uses to determine its actions
  - Can be any algorithm
- Neural network policy
  - Take observations as inputs
  - Output the actions to take
- Policy can have parameters
  - E.g., probability *p* of taking an action
- Policy gradients algorithms:
  - Evaluate the gradients of rewards with regard to policy parameters
  - Follow the gradients towards higher rewards

# AI Learns to Walk



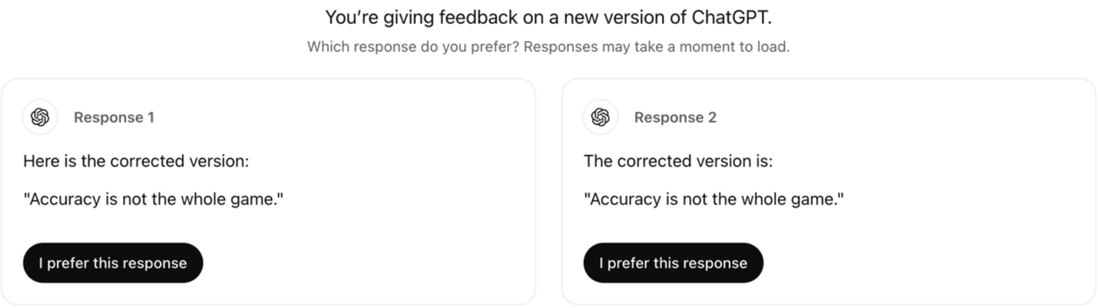
# Why Reinforcement Learning

Notes

- When label is unavailable, use rewards to learn
- Solves sequential decision-making problems where the outcome depends on a sequence of action
- Dynamic or changing environments
  - For example, game-playing or robotics

# Recent Developments

- RLHF: Reinforcement Learning from Human Feedback
- Align an intelligent agent with human preferences
- Use in large language models such as ChatGPT



You're giving feedback on a new version of ChatGPT.

Which response do you prefer? Responses may take a moment to load.

Response 1

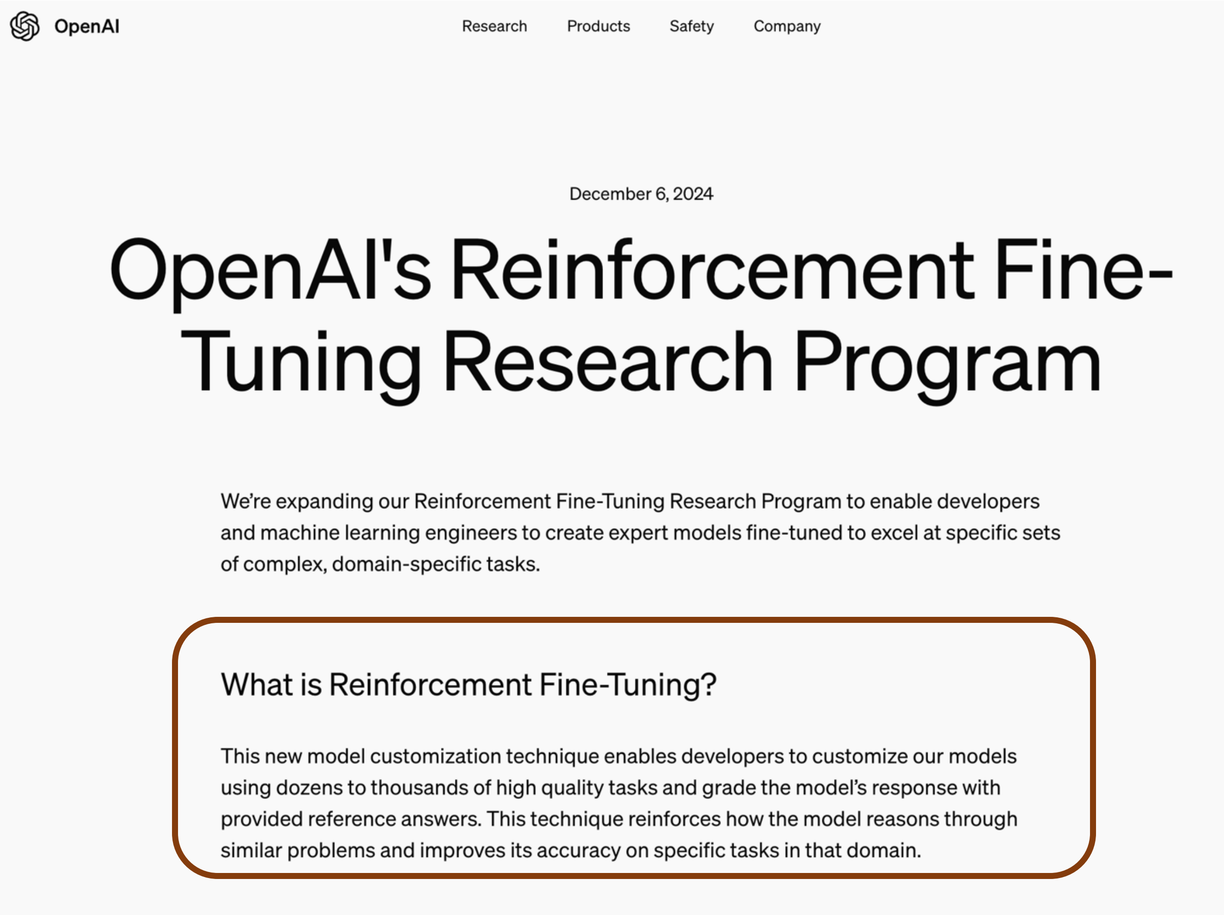
Here is the corrected version:

"Accuracy is not the whole game."

Response 2

The corrected version is:

"Accuracy is not the whole game."



December 6, 2024

# OpenAI's Reinforcement Fine-Tuning Research Program

We're expanding our Reinforcement Fine-Tuning Research Program to enable developers and machine learning engineers to create expert models fine-tuned to excel at specific sets of complex, domain-specific tasks.

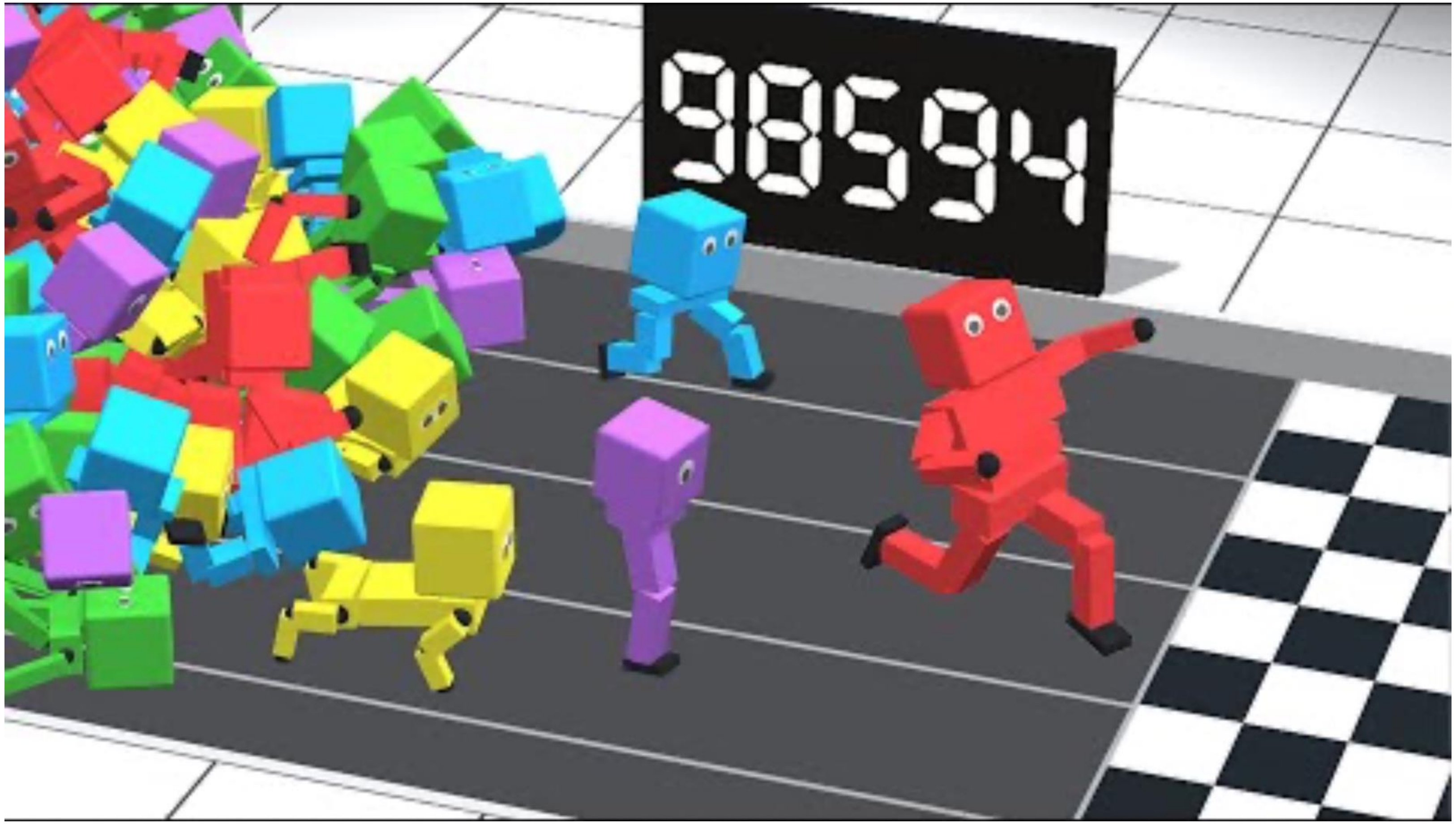
## What is Reinforcement Fine-Tuning?

This new model customization technique enables developers to customize our models using dozens to thousands of high quality tasks and grade the model's response with provided reference answers. This technique reinforces how the model reasons through similar problems and improves its accuracy on specific tasks in that domain.

# Interview Tips

- Not testing, but matching
  - Accuracy score matters, but not the whole game
- No one knows everything about AI/ML
- Showcase your skills
- Do you research: company, team, hiring manager, job description
- A question you can ask yourself: *what type of teammates would you like to work with?*
- If this one does not work, just move on to the next opportunity

# AI Olympics



https://www.youtube.com/watch?v=pJPdW8WWAso&t=1s

*Class Exercise Time!*

# Lab 7 and Simulated Environment

- Reinforcement learning needs an “environment”
- OpenAI Gym https://openai.com/index/openai-gym-beta/
- Handed over to the Farama Foundation: https://farama.org/Announcing-The-Farama-Foundation
- Renamed to Gymnasium: https://gymnasium.farama.org/#
- We will use Cart Pole environment: https://gymnasium.farama.org/environments/classic_control/cart_pole/