

CS 291 – AI Agents

Intro

Prithviraj Ammanabrolu

Logistics

- First day of Class: Today
- Last day of Class: Mar 13
- T/TR 3:30-4:50 CSE 2154
- TA: Bosung Kim
- Website: pearls-lab.github.io/ai-agents-course
 - Office Hours etc are on website
- Class discussions via Piazza (link on Canvas)

What will we cover?

- What are agents? Starting in the 1900s to today
- Simulation environments, how they are built
- Search / Planning in simulations
- Monte Carlo Methods
- Basics of Deep RL
- LLM Basics
- Deep RL for language agents
- Other LLM Agents methods – tool use, prompting
- Human AI Collab Basics
- Agent Safety

What should I know already?

- Intro to AI
 - BFS/DFS/A*
 - Basic Classical RL - Value Iteration
- Intro to ML
 - Perceptron / Classification
- Intro to Deep Learning
 - Huggingface LLM finetuning
 - Forward pass for self attention / basic GPT

HW 0

- We have a HW that we released
- Most of it is autograded, submit code + things mentioned in the readmes.
- If you don't get $\geq 70\%$ within a couple of week (1/21) by *yourself* (be honest!), the class will be difficult for you
- Use how much time you spend as estimate of what you'll spend in this class per week

Grading

- 40% - 4 Homeworks (including HW 0) - to be done individually
- 40% - Final Project – to be in groups of ~5
- 20% - Class participation

Class Participation

- Sign up with your project group to do paper presentations starting 1/21
- Paper topics will be related to the previous lecture and happen at the end of class for ~20 min + discussion of 10+ min
- Earlier presentations will be graded more leniently than others
- Send us a draft of the slides a couple of days ahead