# ADVANCED REINFORCEMENT LEARNING

**Summer 2023** | **Instructors:** Pulkit Agrawal, Cathy Wu | professional.mit.edu/arl

---

**Note:** All times are US Eastern Daylight Time. Schedule is subject to change.

## DAY 1 (9:00am–6:00pm)

<table>
<thead>
<tr>
<th>Time</th>
<th>Session Description</th>
</tr>
</thead>
</table>
| 9:00–9:30 AM  | **WELCOME**  
- Meet and greet  
- Problem description from class attendees |
| 9:30–11:00 AM | **SESSION 1: OVERVIEW OF DEEP RL AND ITS LIMITATIONS**  
- Summary of popular algorithms like DQN, PPO, A3C  
- Motivation for advanced topics  
  - Guidance from experts: Learning from demonstrations  
  - Use of existing datasets: Offline RL  
  - Increasing data efficiency via model learning  
  - Multi-task RL  
  - Curriculum learning  
  - Advanced exploration and exploitation |
| 11:00–11:30 AM | **BREAK** |
| 11:30 AM–1:00 PM | **SESSION 2: STATE-OF-THE ART POLICY GRADIENT ALGORITHMS**  
- Advantage Actor Critic (A2C)  
- Advantage Actor Critic (A2C)  
- Trust Region Policy Optimization (TRPO) Proximal Policy Gradients (PPO)  
- Hyperparameters and tricks in Policy Gradients  
  - How deep should my network be? What learning rate should I use? How should the network be initialized? |
| 1:00–2:00 PM | **LUNCH** |
| 2:00–3:00 PM | **SESSION 3: INTERACTIVE SESSION (EMPIRICAL RIGOR)**  
- Go through Jupyter Notebook with code  
- Run algorithms and modify them to analyze what matters  
- PPO and implementation matters  

**ADVANCED Q-LEARNING**  
- Prioritized Experience replay  
- RAINBOW: Combining several improvements in Deep Q-Learning  
- Distributional RL |
| 3:00–3:30 PM | **BREAK** |
| 3:30–5:00 PM | **SESSION 5: LEARNING FROM DEMONSTRATIONS**  
- Learning from experts  
  - Behavior cloning  
  - DAGGER  
  - Augmenting Behavior Cloning with RL  
  - Observational learning  
- Practical issues and solutions: Causal confusion, inability to go beyond expert performance, etc. |
| 5:00–6:00 PM | **RECEPTION** |
## ADVANCED REINFORCEMENT LEARNING

**Summer 2023 | Instructors: Pulkit Agrawal, Cathy Wu | professional.mit.edu/arl**

*Note: All times are US Eastern Daylight Time. Schedule is subject to change.*

### DAY 2 (9:00am–5:00pm)

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
</table>
| 9:00–10:00 AM | **SESSION 1: OFFLINE RL**  
- Introduction to Offline RL  
- Off-policy evaluation + Applications  
- State-of-the-art in Offline RL  
- Applications of Offline RL |
| 10:00–11:00 AM | **SESSION 2: CHATGPT AND ALIGNMENT USING RL**  
- How RL uses human feedback to improve performance of chatbots  
- Brief discussion on aligning AI with human values |
| 11:00–11:30 AM | **BREAK** |
| 11:30 AM–12:30 PM | **SESSION 3: INTERACTIVE SESSION (SIM2REAL)** |
| 12:30–1:30 PM | **LUNCH** |
| 1:30–2:15 PM | **SESSION 4: MONTE CARLO TREE SEARCH**  
- Applications to go  
- Applications to other areas |
| 2:15–3:00 PM | **SESSION 5: MULTI-AGENT RL** |
| 3:00–3:30 PM | **BREAK** |
| 3:30–4:30 PM | **SESSION 6: MODEL BASED AND GOAL BASED RL** |
| 4:30–5:00 PM | **SESSION 7: AMA** |

### ADDITIONAL SESSIONS

Three additional 60-minute sessions will be held after the class to serve as a problem clinic.