# Advanced Topics in Reinforcement Learning

## Lecture 23: Empirical Design and Ethical Considerations

Josiah Hanna
University of Wisconsin — Madison

# Announcements

- Next week: Offline RL
- No class Thursday! 🎉

# Learning Outcomes

After today, you will be able to:

1. Analyze strengths and weaknesses in RL empirical design.

2. Identify and discuss ethical considerations that arise in RL.

# Kunzhao’s Presentation

- Deep Reinforcement Learning that Matters
- Henderson et al. 2018.
- Slides

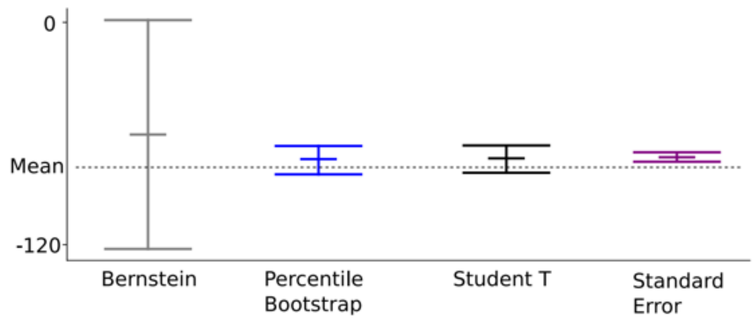
# Common Random Seeds

- You plan to compare algorithms A and B in simulated environment X.
- Since data in RL is stochastic, what if algorithm A gets unlucky with observed data?
- To prevent this, fix a set of common random seeds. Ensure that each algorithm is ran once with each seed.

```python
import random

seeds = [random.randint(1e6) for _ in range(n_trials)]

for seed in seeds:
    # Run algorithm A with seed
    ...

    # Run algorithm B with seed
    ...
```

# How many runs / seeds?

- Claims should match evidence presented.
  - Example: Task A can be solved by Algorithm X. How many seeds?
  - Example: Algorithm X has higher mean return at convergence compared to Algorithm Y on Task A. How many seeds?
- What if you don't have the compute budget to run sufficient seeds?
- Which confidence interval to use?
- Don't fix seeds for experiments.




**Note:** Likely many Python packages for different CIs, e.g., Seaborn

# Which Baselines?

- The purpose of baselines is to contextualize the numbers that a paper reports to support its claims.
- Possible baselines:
  - Alternative approaches (most common type in ML papers)
  - Lower bounds (e.g., random agent)
  - Upper bounds (e.g., an algorithm with priviledged knowledge, humans, etc.)

# Hyper-parameter Selection

- Report how hyper-parameters selected. Why?
  - Best hyper-parameters are problem dependent and so the process is more important than the final hyper-parameter values.
- How to select:
  - Graduate student descent
  - Random search
  - Grid search
  - More advanced: Bayesian optimization, population-based training, neural architecture search.

# Community Standards

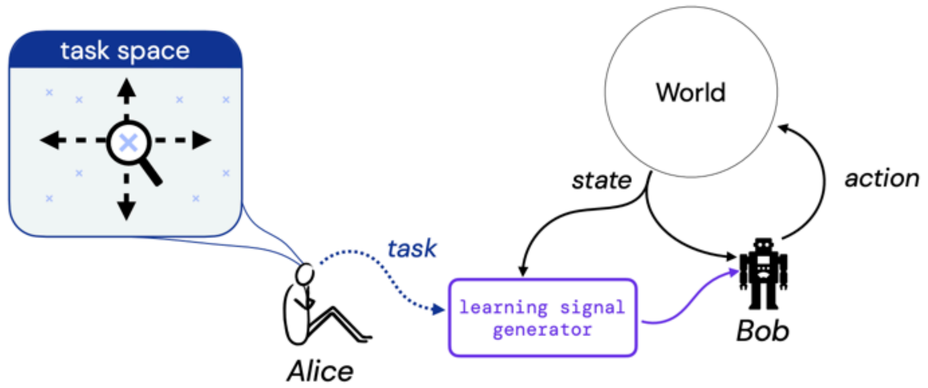
- What should conference and journal standards for reproducibility be?
  - Reproducibility challenges
  - Reproducibility check-lists
- Culture challenges:
  - Emphasis on positive results and novelty.
  - Lack of precision about claims, e.g., “our algorithm dominates the baseline”
  - Low benefit for reproducing work of others and considering incremental questions.
  - Often tension between mathematically well-founded approaches and strong empirical performance.

# Sharing Code

- Is code sharing sufficient?
  - Not necessarily, may require missing compute and data resources.
- Why do different code-bases give different performances?
  - Deep RL implementations often have hidden tricks that go beyond the base algorithm.
  - Example: observation normalization, i.e., divide observed state variables by a running average of their standard deviation.

# Ethical Issues in RL

- Standard ML issues around bias and fairness persist.
- The reward function defines the goal of a task…
  - Insufficiency of Markov reward formalism (Abel et al. 2021)
  - Risk of misalignment
- .. not how to achieve it.
  - Need to consider side-effects of optimization.
  - Example: Scheduling algorithms disadvantage certain classes of  shift workers.




Abel et al. 2021

# Mahesh’s Presentation

- On Targeted Manipulation and Deception when Optimizing LLMs for User Feedback
- Williams et al. 2025.
- Slides

# Summary

- Deep RL experiments are very stochastic. Makes evaluation of deep RL algorithms a challenge.
- Careful empirical design is necessary to support the claims made.
- RL optimizes a given reward function — one challenge is to align the reward and resulting behavior with human values.

# Action Items

- Offline RL reading for next week. Last reading!
- Continue making progress on your final project.