Nice to meet you!

- Chicheng Zhang
- Assistant Professor at CS @ UA (since 2019)
- Research interests: algorithms & theory for learning to make sequential decisions
- Email: chichengz at cs.arizona.edu

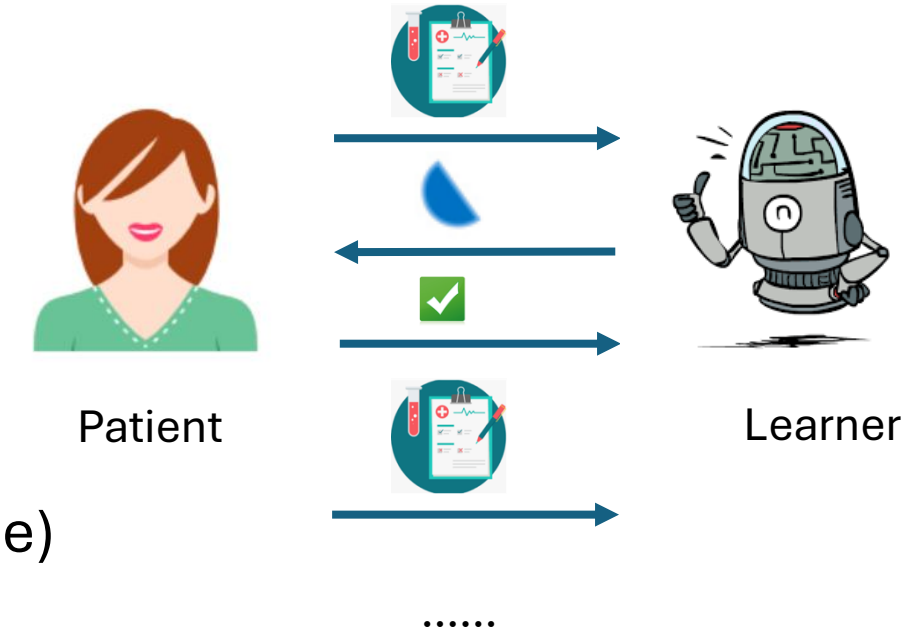
Imitation Learning

Motivating example: medical treatment

A ‘robot doctor’ is paired with a patient to provide personalized medical consultations

Repeatedly:

- Receives state (medical history)
- Takes an action (prescription)
- Receives reward (the patient feels better / worse)
- Transitions to new state



- How can the robot doctor help keep the patient healthy in the long run?

Applications of sequential decision making (control)

- Robot arm manipulation

- Game playing

- Conversational assistant
  - (e.g. Reinforcement Learning from Human Feedback in training ChatGPT)

- Autonomous driving

- …

Imitation learning (IL)

Imitation learner

Control Policy

```
If dist_obstacle < 5m:
        brake
Else:
        accelerate
```

- Applications:
  - Autonomous driving
  - Robot control
  - Game playing

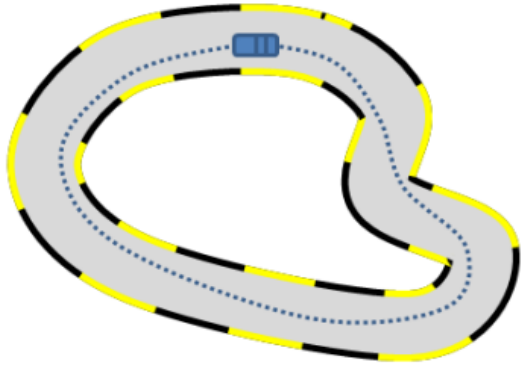
Modern imitation learning systems

- The ALOHA Robot (Stanford, 2024): Trained using imitation learning, expert demonstration data collected by teleoperation

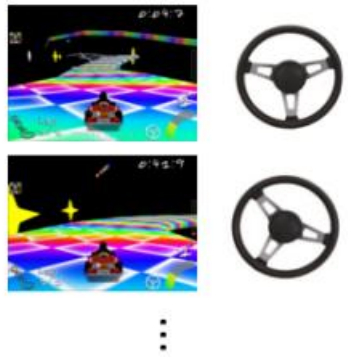
Example: learning to drive from demonstrations

Training stage:

Expert Trajectories



Dataset



Summarize rule from data

Policy

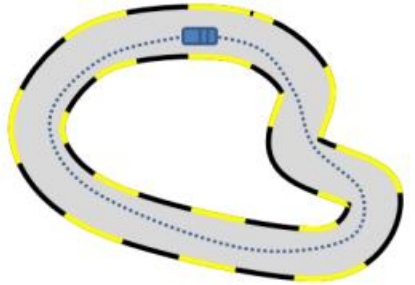
```
If dist_obstacle < 5m:
        brake
Else:
        accelerate
```

Test stage:

Policy

```
If dist_obstacle < 5m:
        brake
Else:
        accelerate
```

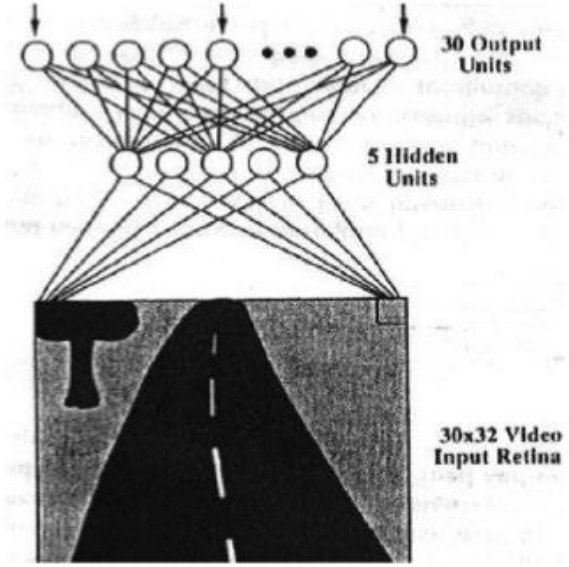
Deployment in environment



(Images from Stephane Ross's slides)

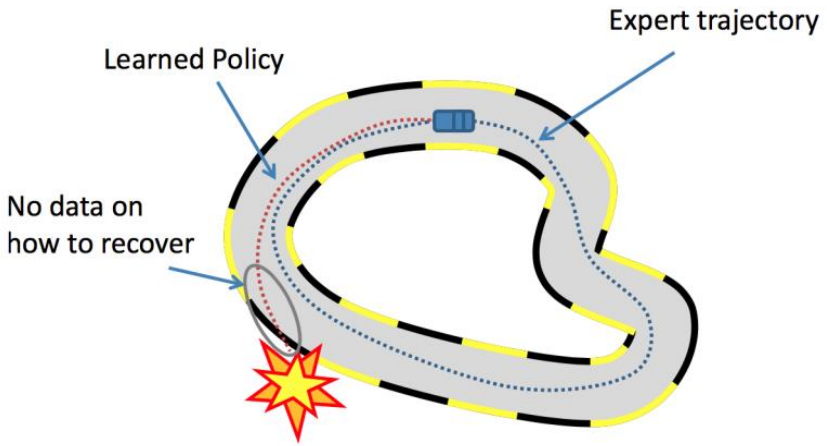
Early experiments of Imitation Learning

- ALVINN: Autonomous Land Vehicle in a Neural Network (Pomerleau, 1989)



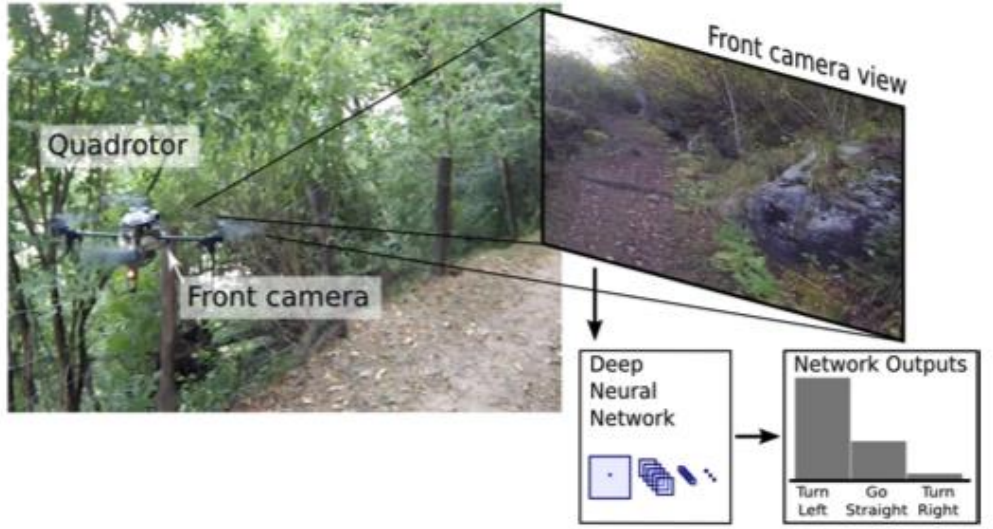
The cascading error problem

- Mistake -> unseen observations -> mistakes again



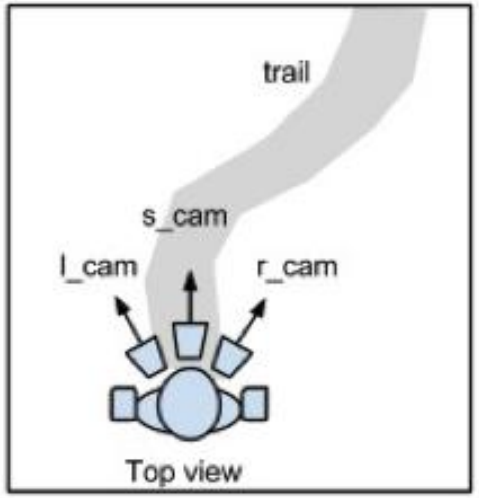
- Bojarski et al, 2016: "*Training with data from only the human driver is not sufficient. The network must learn how to recover from mistakes*."

How to learn to recover?

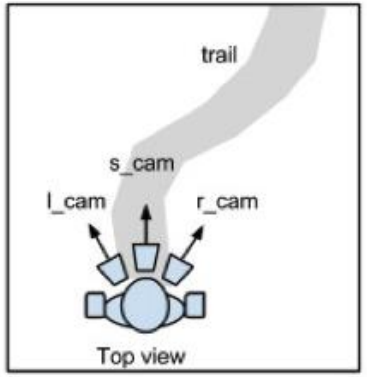


Idea 1: (Giusti et al, 2015)

- Goal: train a drone that can follow trails
- Key idea: ask a hiker to wear 3 on-head cameras & walk along trail



Observations seen by the cameras



Observations seen by left cameras

Action: turn right

Observations seen by right cameras

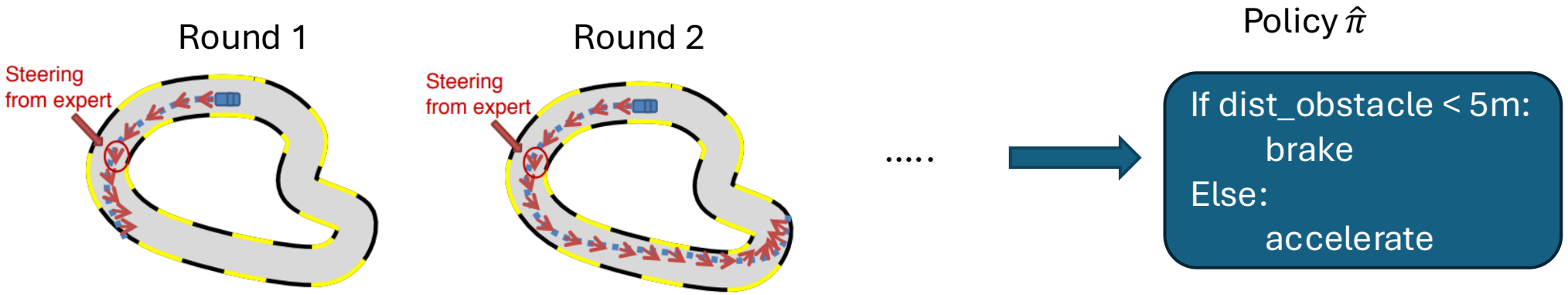
Action: turn left

Observations seen by straight cameras

Action: go straight ahead

How to learn to recover?

- Idea 2: leverage interaction with the human expert
- Interactive imitation learning: learner adaptively *queries* expert for demonstrations

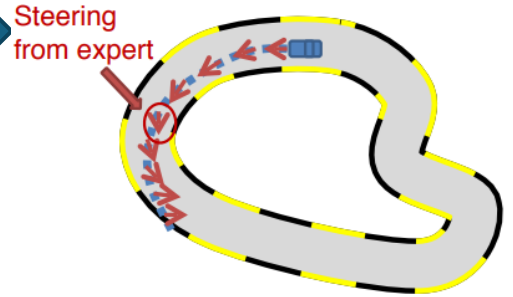


- Why is this effective?
  - The learner receives *targeted* corrective feedback, and thus can learn to recover from its own mistakes

Dataset Aggregation (DAgger)

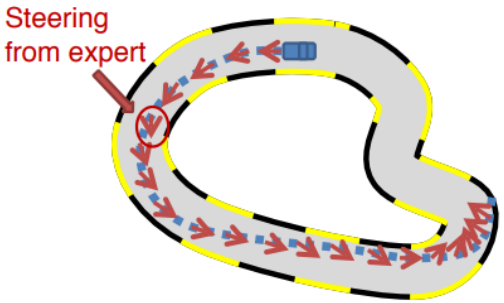
- Goal: learn a driving policy that can mimic the human expert
- Repeatedly:

Learned policy from historical demonstration data



New demonstration data

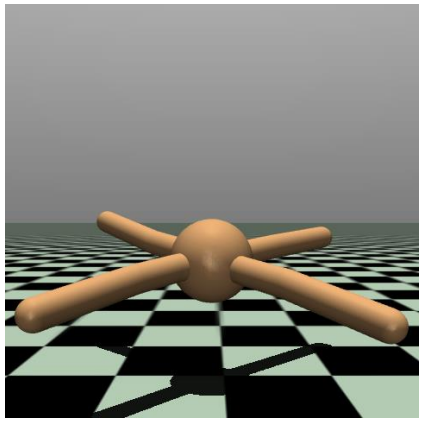
Learned policy from (updated) historical demonstration data



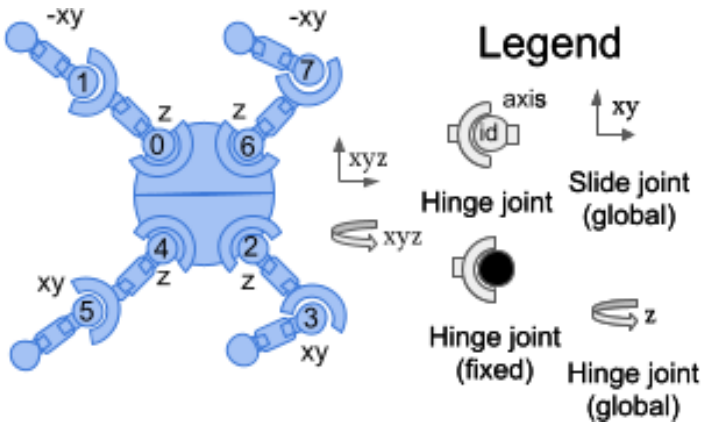
New demonstration data

- Return final policy learned from *all* historical demonstration data

Dagger is effective in robotic control

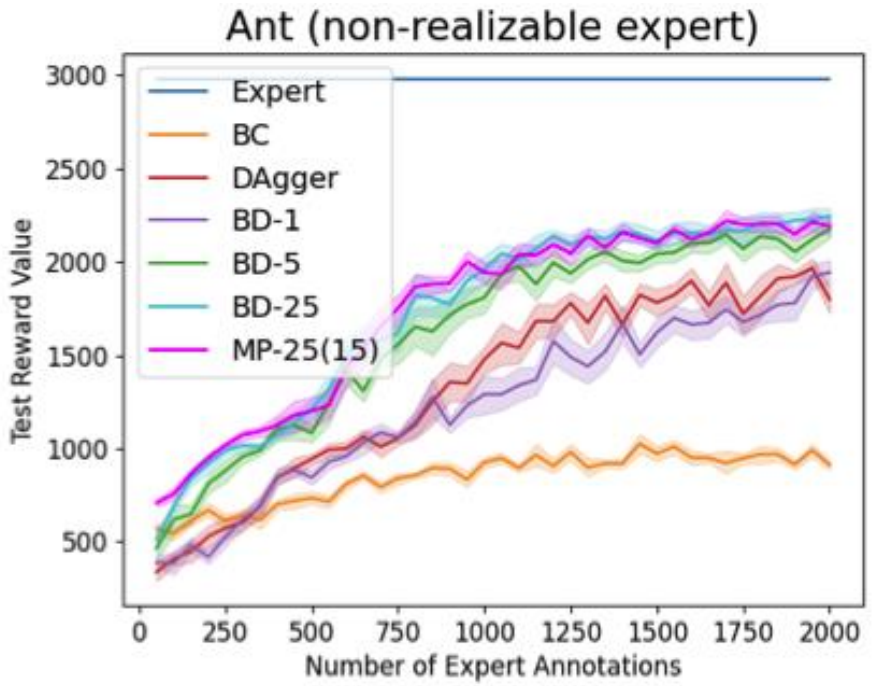


- Example control task: Ant
- Goal: Move the Ant robot forward
- States: 105-dimensional vectors
  (pose, joint angles and velocities)
- Actions: 8-dimensional vector (torques on joints)



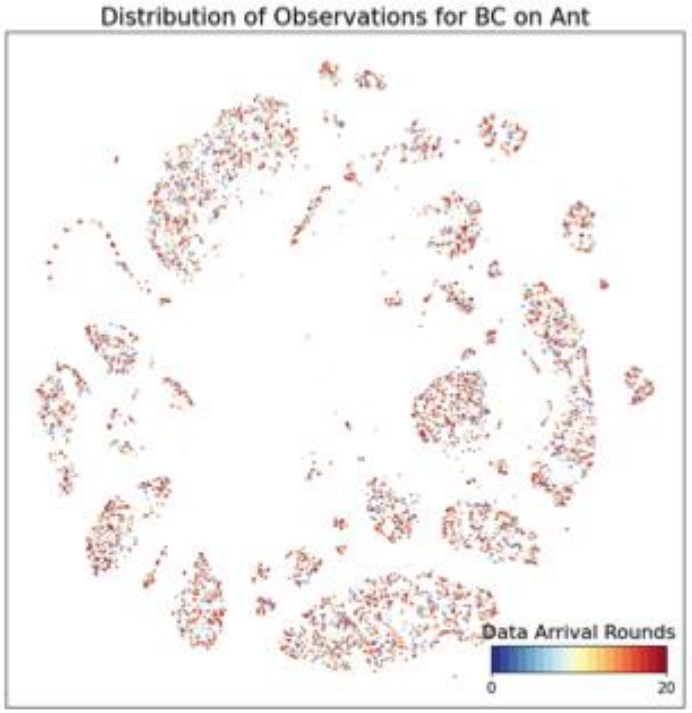
DAgger is effective in robotic control

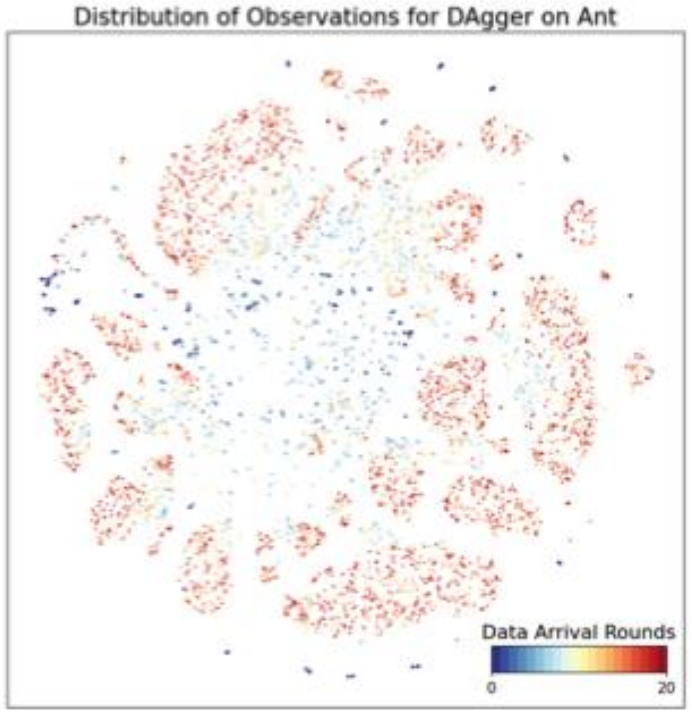
- Join work with Yichen Li (PhD Graduate 2025, now at Amazon):
- x-axis: number of the expert annotations
- y-axis: performance of trained agent (higher the better)
- Behavior Cloning (BC): offline expert data
- All other lines are variants of DAgger



DAgger collects a broader coverage of data

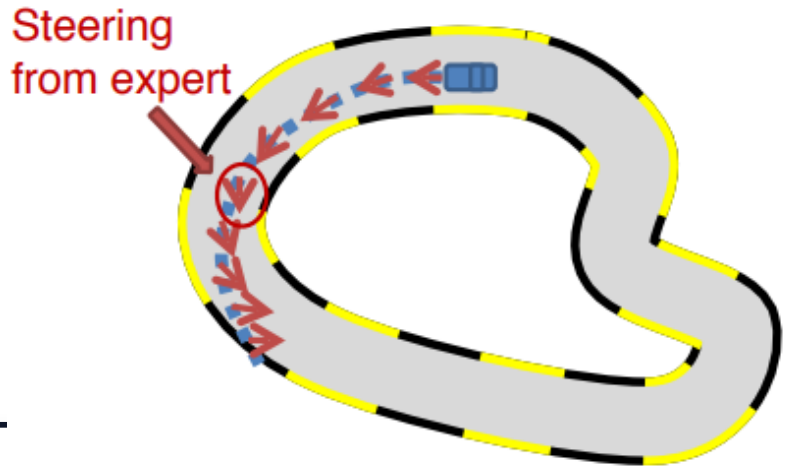
We visualize the observations collected by offline expert demonstration and DAgger by projecting them to two dimensions





Sample efficiency is important

- It takes a lot of effort for the expert to provide steering demonstrations



- More demonstrations = more $$$

Data Annotation Market Growth Projections

Source: Market research 2026

$2.32B	$9.78B	33.27%	80%
Market Size 2025	Projected Size 2030	Annual Growth Rate (CAGR)	Of ML Budget for Annotation

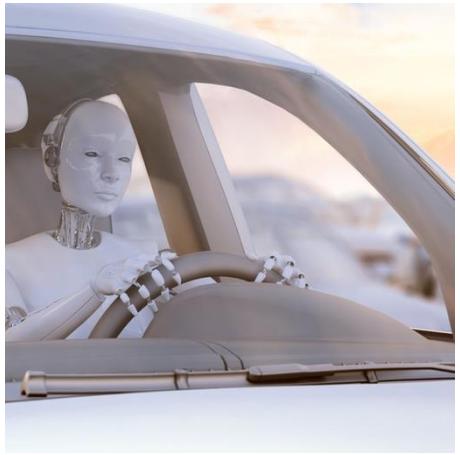
secondtalent.com

- How can we make our learning cost-efficient?

Making DAgger more sample-efficient [Li & Zhang, 2025]

We can just collect one expert annotation per learning round!

- Repeatedly:



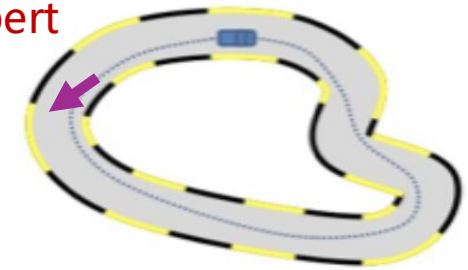
Learned policy from historical demonstration data

One additional expert demonstration

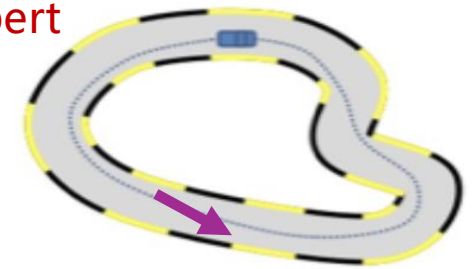
Learned policy from historical demonstration data

One additional expert demonstration

Steering from expert



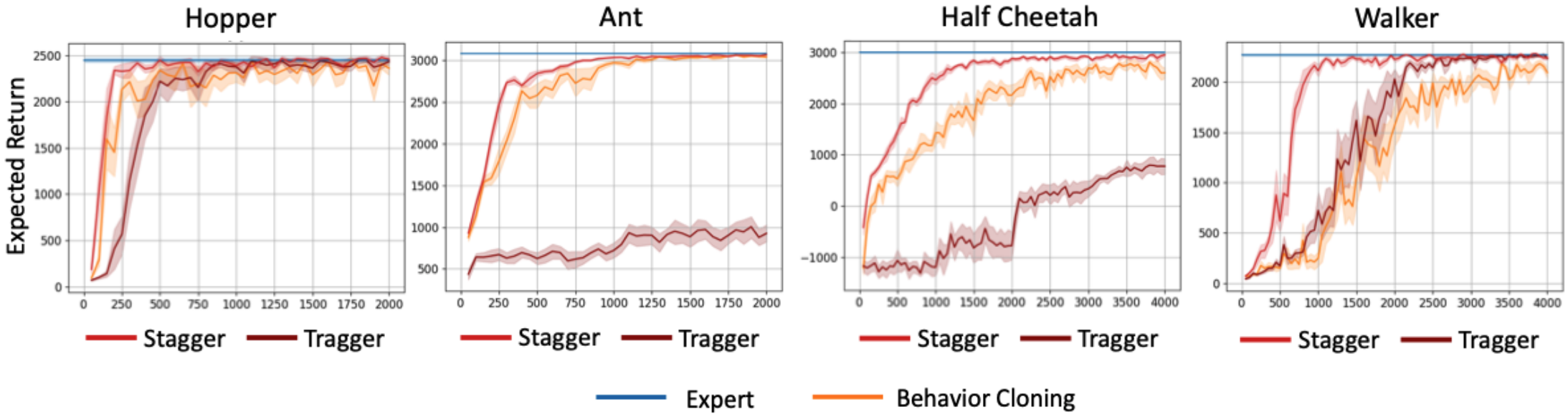
Steering from expert



- Return final policy learned from historical demonstration data

Stagger -- State-wise DAgger [Li & Zhang, 2025]

- When measuring the sample cost using the number of states annotated, Stagger learns much faster than
  - Behavior Cloning
  - Tragger (DAgger with one trajectory per iteration)

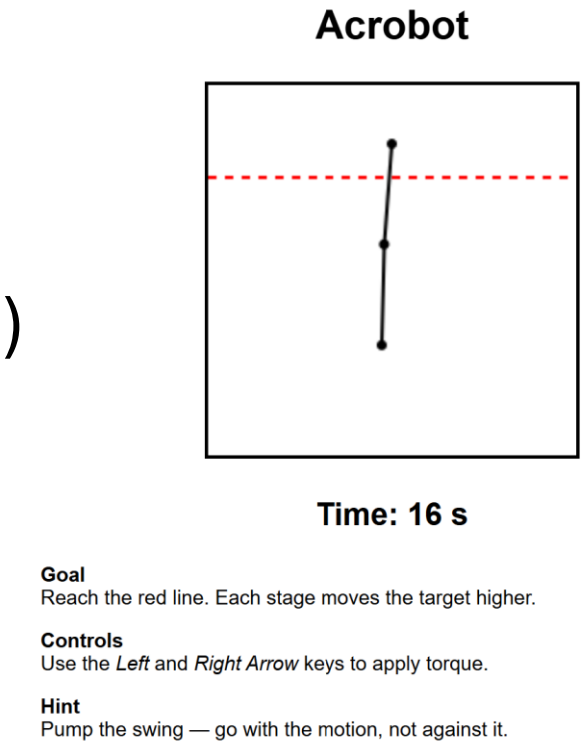


Research opportunities

- We have a web-based portal for imitation learning data collection
  https://imitation-learning.onrender.com
  (we will have you as the demonstrating expert)

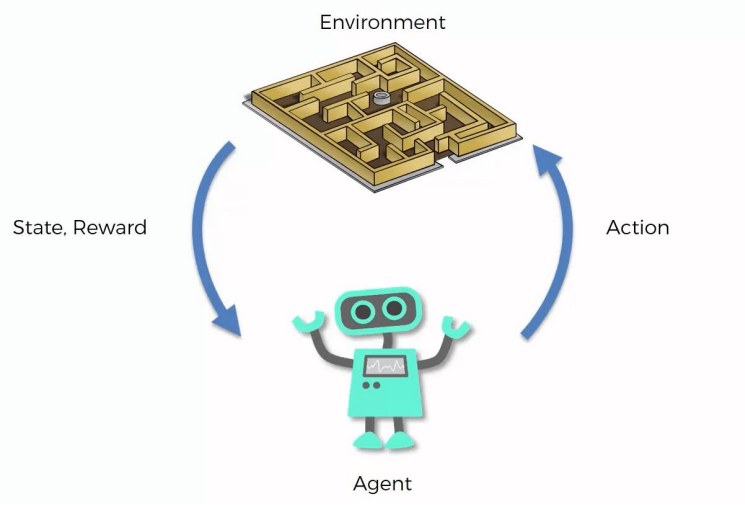
(thanks to Elise Bushra, undergraduate thesis mentee)

- If you are interested in imitation learning, you are welcome to reach out



Closing remarks

- Learning to make sequential decisions exposes the hard parts of intelligence: acting, exploring, planning



- Learning from human interaction: synergize the computational power of machines and smart insights of humans