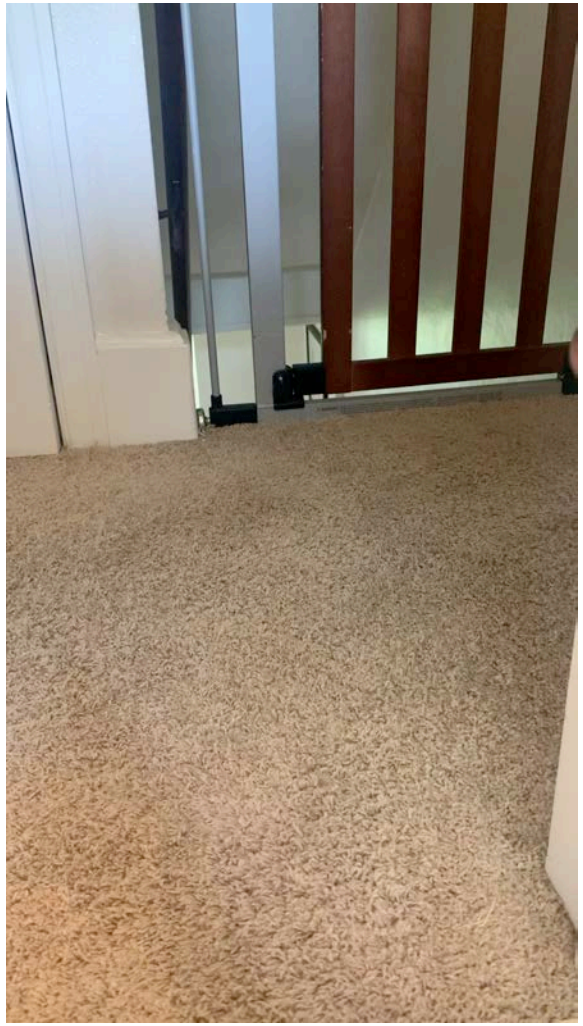
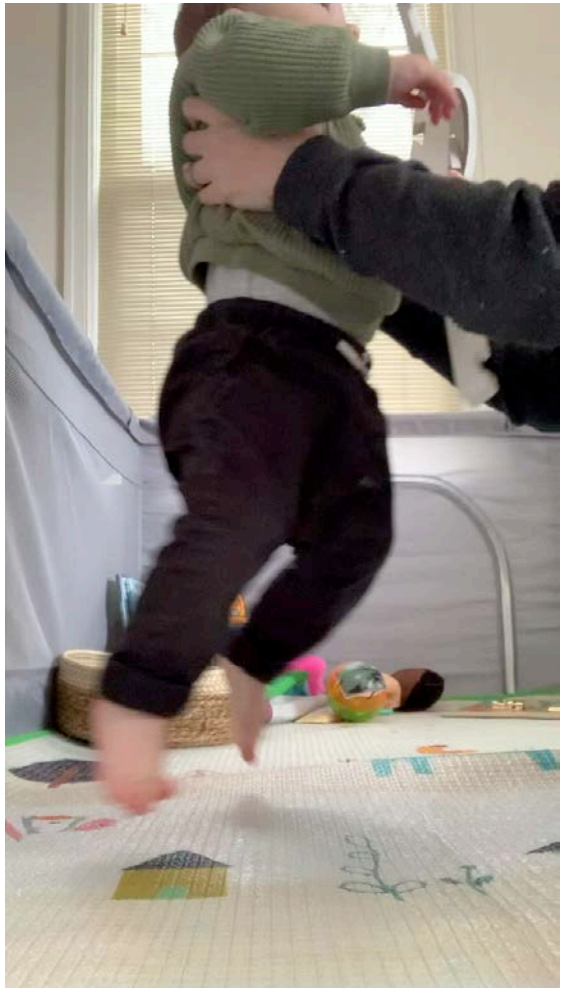
Artificial Intelligence from Interaction and Reward

Josiah Hanna
Computer Sciences
Wisconsin — Madison



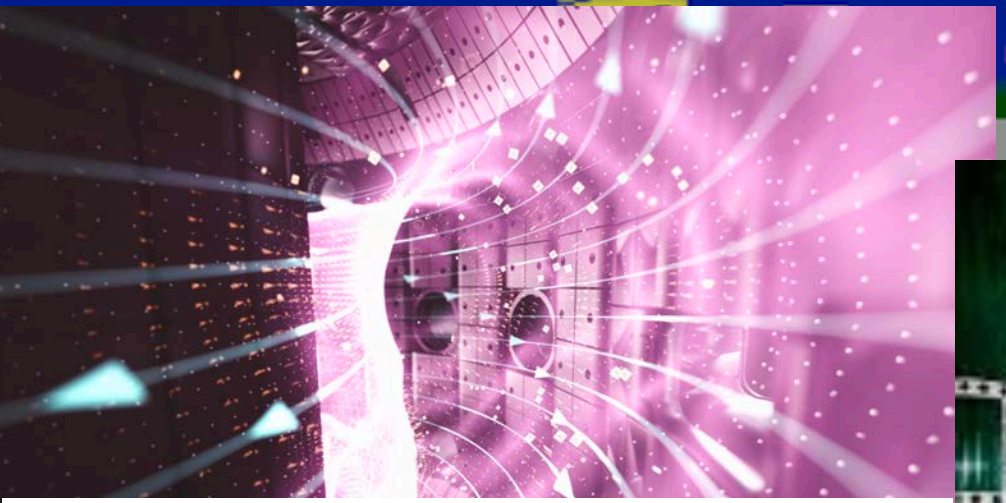


Hanna and Stone, 2017. Hanna et al. 2021.

Algorithms that enable computers and robots to exhibit the same adaptive, flexible, and goal-directed behavior that we observe in people and animals.


900

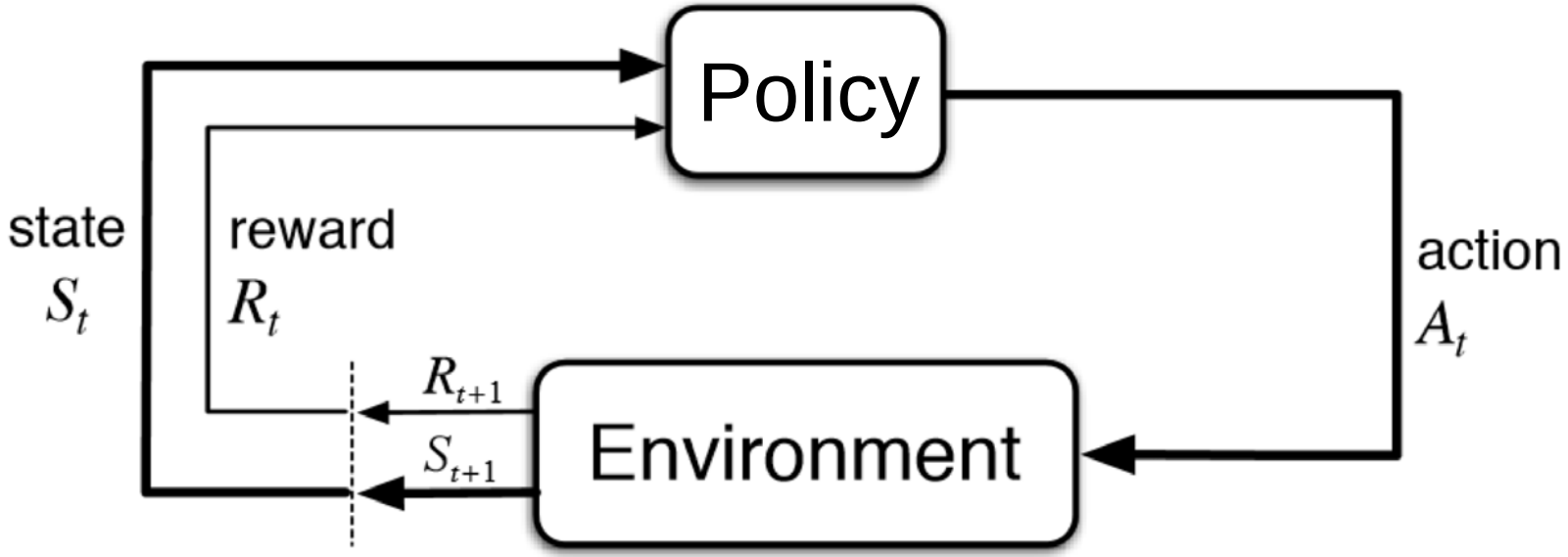

ALPHAGO
00:10:29
AlphaGo
LEE SEDOL
00:01:00




INTERNET
ADVERTISING
ENGINE
WEB
PER
CONSUMER
ONLINE
MARKETERS
PRODUCTS
BUSINESS
PRODUCT
CONSUMERS
SPECIFIC
MARKETING
SEARCH

ChatGPT

The Reinforcement Learning World



$$\underbrace{S_0, A_0, R_0, S_1, \ldots, S_L, A_L, R_L}_{\texttt{Trajectory}}$$

Find the policy that maximizes the expected sum of rewards per trajectory.

Why is interactive learning hard?

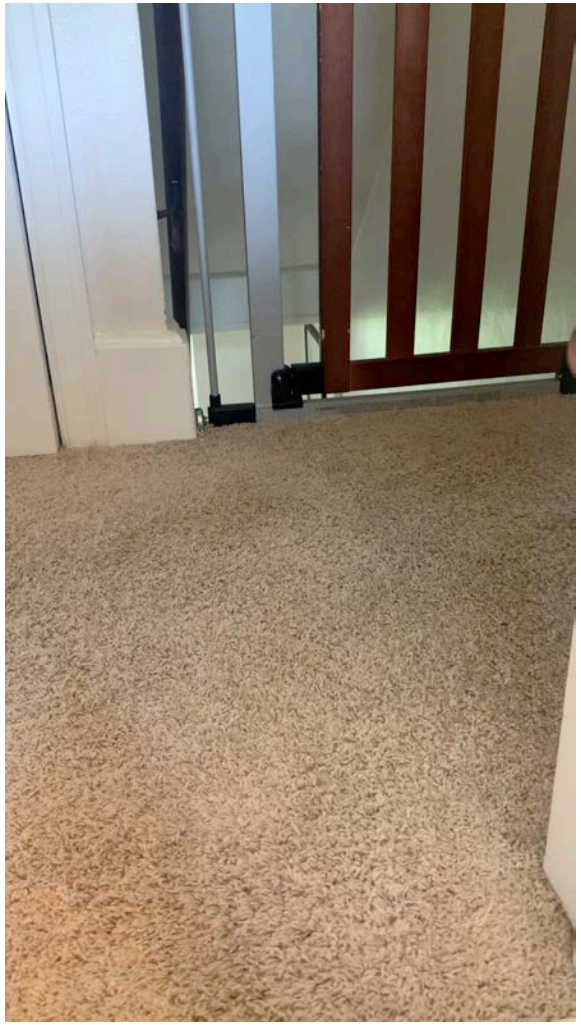
- Exploration vs. Exploitation
    - Should you keep trying actions that led to reward in the past or try new actions that might lead to even more reward in the future?
    - Example: when you go out to eat at a familiar restaurant, do you choose an option you enjoyed before or try something new to maybe find something even better?

Why is interactive learning hard?

- Credit Assignment:
    - May take many actions before reward is received. Which ones were most important?
    - Example: you study 15 minutes a day all semester. The morning of the final exam, you eat a bowl of yogurt. You receive an A on the final. Was it the studying or the yogurt that led to the A?
    - Not trivial for people and animals either — done poorly leads to superstitious behavior!

Reinforcement Learning in Practice

- Today, many AI systems learn from large, fixed datasets ==> No interaction.
- Offline Reinforcement Learning ==> learn from data produced by some other agent.
- Sim2Real RL ==> train in a simulation then deploy in the real world.

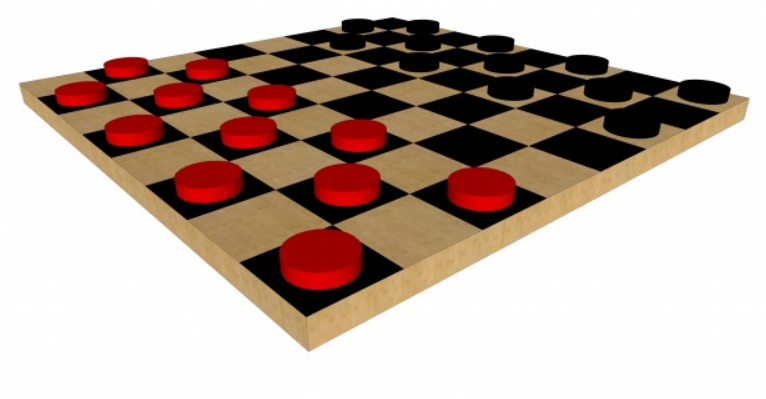


My Research

Find domains where RL has potential, identify challenges for the application of RL, and then find solutions to those challenges.

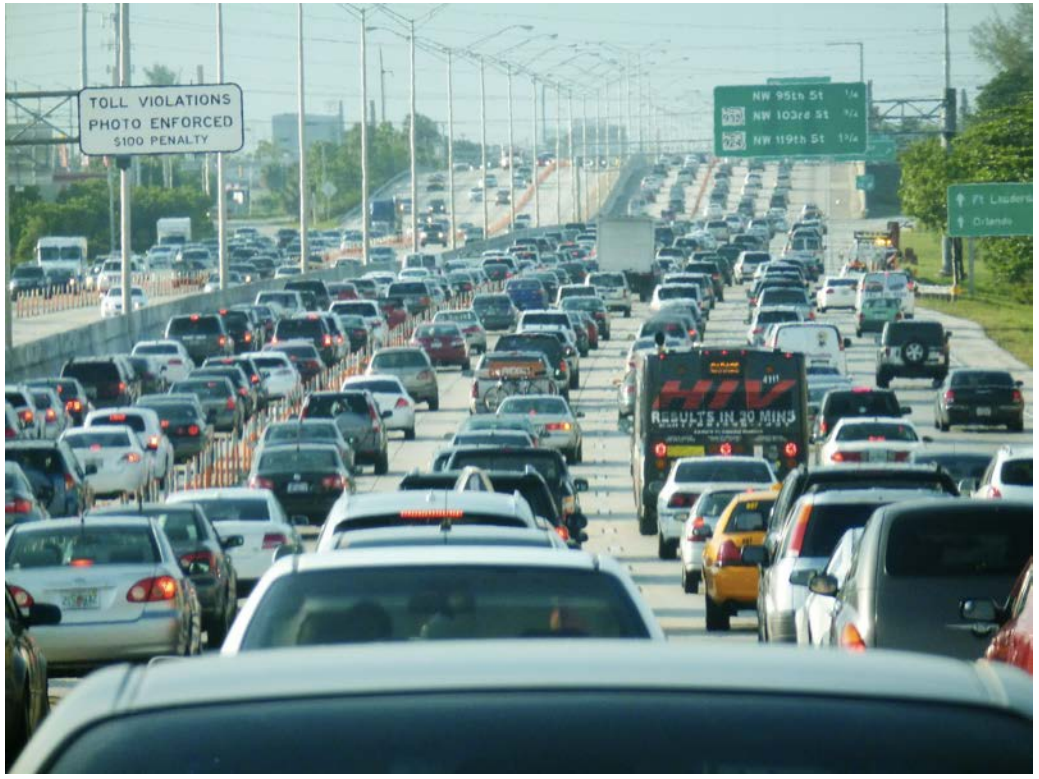
Meeting these challenges is a path towards computers and robots that can learn from interaction and reward.

Artificial Intelligence and Games



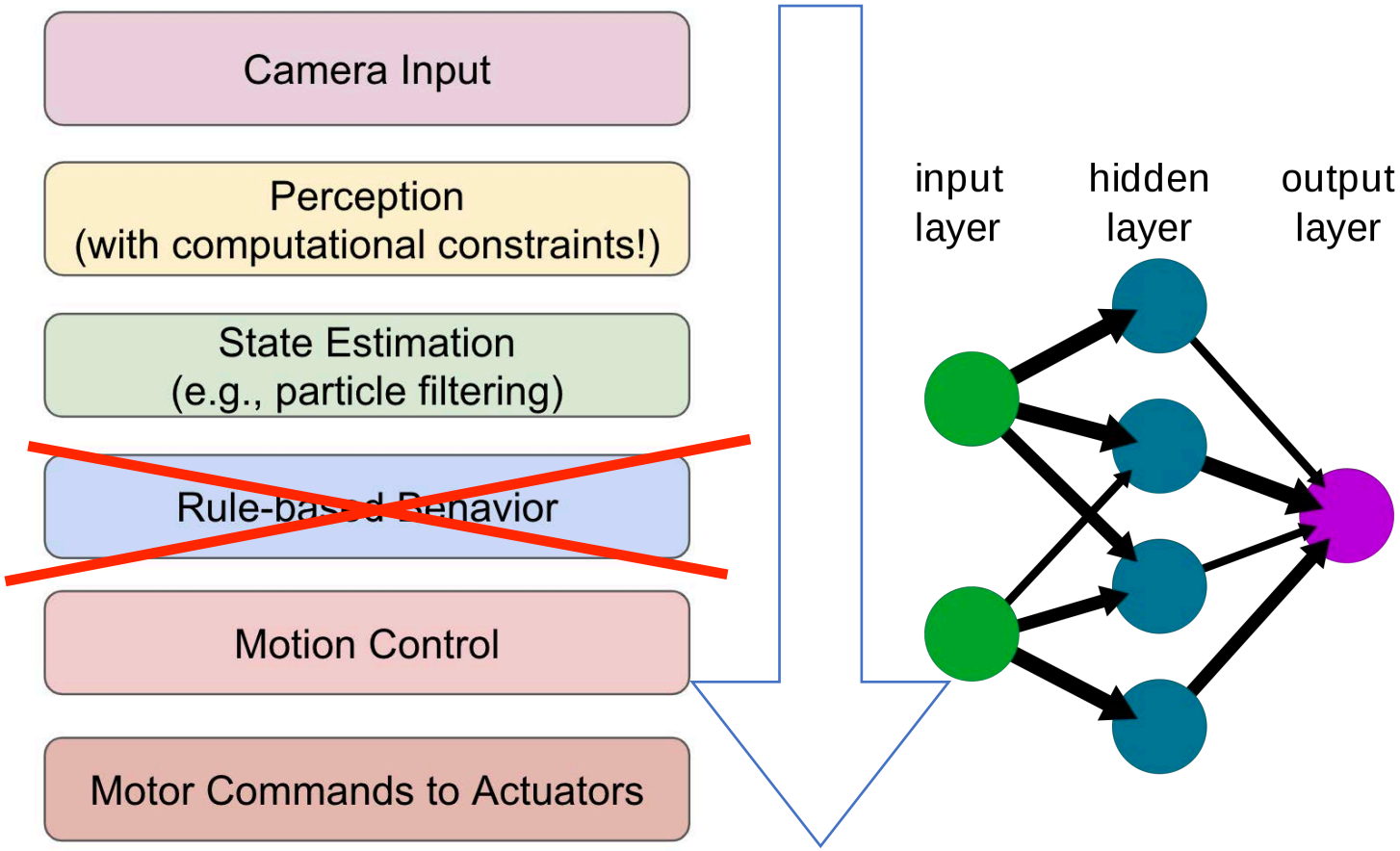
Soccer as a Robotics Challenge

Photo Credit: Wall Street Journal



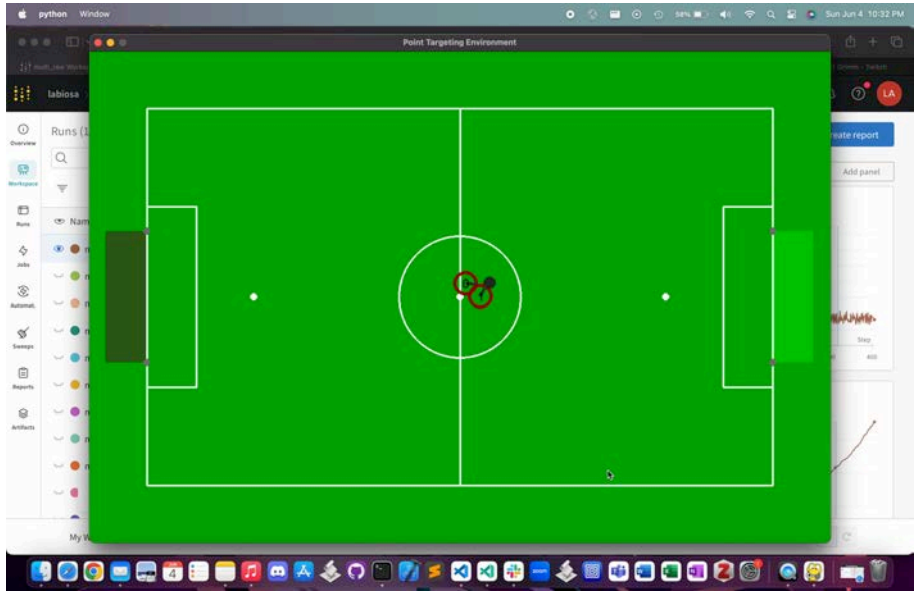
"By the middle of the 21st century, a team of fully autonomous humanoid robot soccer players shall win a soccer game, complying with the official rules of FIFA, against the winner of the most recent World Cup." - Vision statement of the RoboCup Federation

Reinforcement Learning for Robot Soccer



With John Balis, Adam Labiosa, Ben Hong, Chen Li

Reinforcement Learning for Robot Soccer



Train in simulator, deploy to the real robot.

With John Balis, Adam Labiosa, Ben Hong, Chen Li

Reinforcement Learning for Robot Soccer

With John Balis, Adam Labiosa, Ben Hong, Chen Li

Challenging use-cases lead to fundamental research with wide-spread applicability

1. Research Challenge #1: Reinforcement learning algorithms must be able to learn from small amounts of interaction.

2. Research Challenge #2: Can we predict when a learned behavior is safe enough to be deployed as part of a real system?

Challenging use-cases lead to fundamental research with wide-spread applicability

1. Research Challenge #1: Reinforcement learning algorithms must be able to learn from small amounts of interaction.

2. Research Challenge #2: Can we predict when a learned behavior is safe enough to be deployed as part of a real system?

Validating Learned Behaviors

Can we determine the expected performance of a learned policy **before** it is deployed and its actions have real world consequences?

Our work: use historical data on actions and their effects in different states to evaluate an untested, new policy.

Autonomous Driving

Driving requires reasoning about what you cannot see.

Our work: use the behavior of other vehicles to inform us about what we cannot see.

Tolling in Road Networks

Dynamic tolling is being looked to as a means to reduce congestion in road networks. How to update tolls in real time to minimize congestion?

(a) ETC system

(b) ETC gantry

Woodsville Tunnel (71)

Cars/Light Goods/Taxis (Weekdays) ▼

07:00 - 07:30	$0.00
07:30 - 08:30	$0.50
08:30 - 08:35	$1.00
08:35 - 08:55	$1.50
08:55 - 09:00	$1.00
09:00 - 09:30	$0.50
09:30 - 22:30	$0.00

(c) ETC rates

Our work: use the behavior of other vehicles to inform us about what we cannot see.

Challenging use-cases lead to fundamental research with wide-spread applicability

Algorithms that enable computers and robots to exhibit the same adaptive, flexible, and goal-directed behavior that we observe in people and animals.

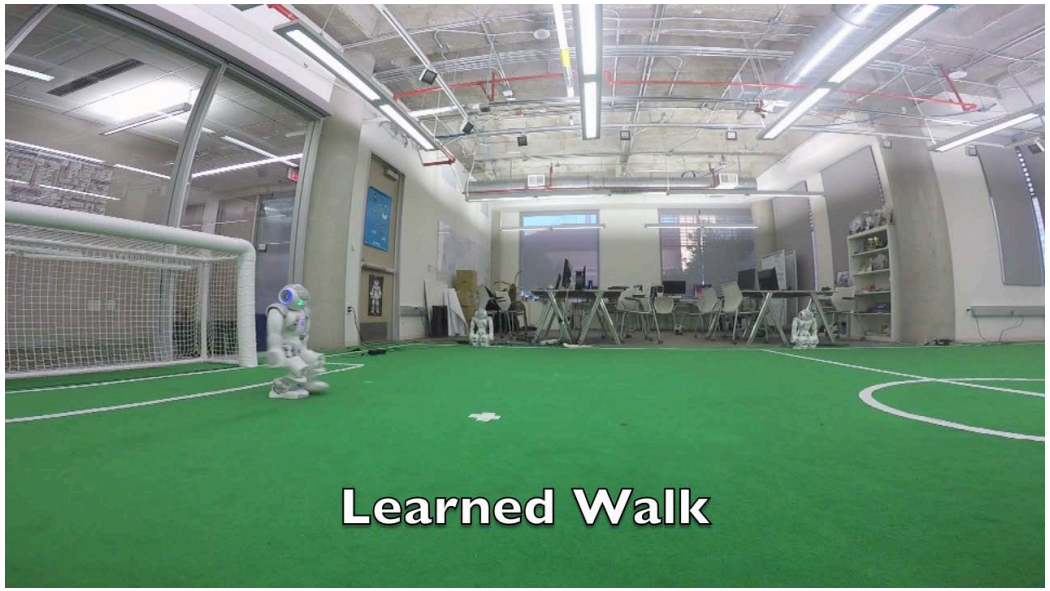
Brahma Pavse

John Balis

Adam Labiosa

Ben Hong

Not pictured but much appreciated: Chen Li



jphanna@cs.wisc.edu

pages.cs.wisc.edu/~jphanna

WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON