UNIVERSITY OF ALBERTA

The Intelligent Robot Learning Laboratory

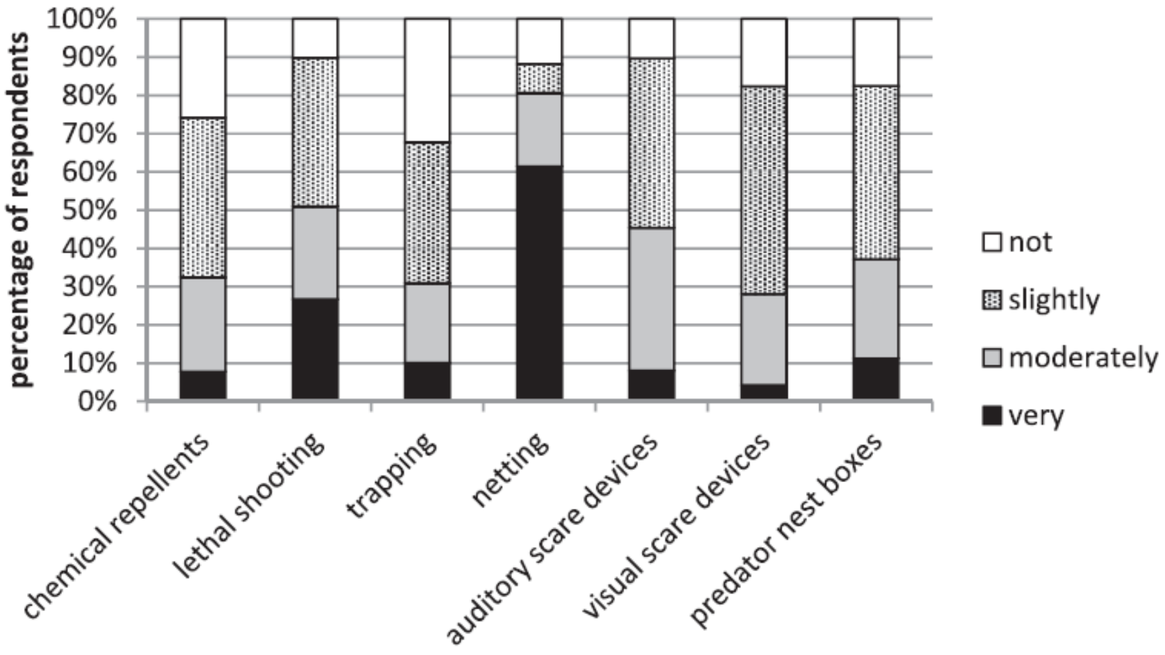
Matthew E. Taylor (Matt)

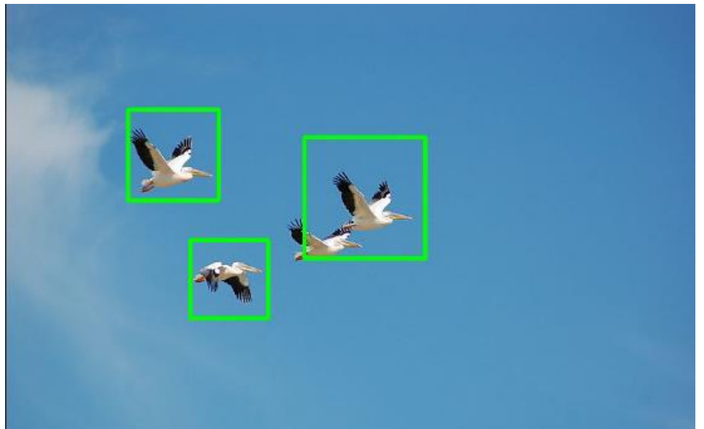
Reinforcement Learning in the Real World:
Challenges and Opportunities for Human Collaboration

air
AI Redefined

Bird Deterrence

- High-value agriculture
- Annual bird damage in WA ~$80 million!
- Low-tech approaches
- We measured bird damage with and without UAVs





Bin Dog

- Controlling a novel “bin-dog” robot to autonomously traverse an apple orchard, ~800 lbs
- Coordinate with human pickers
- Mixed-reality testing

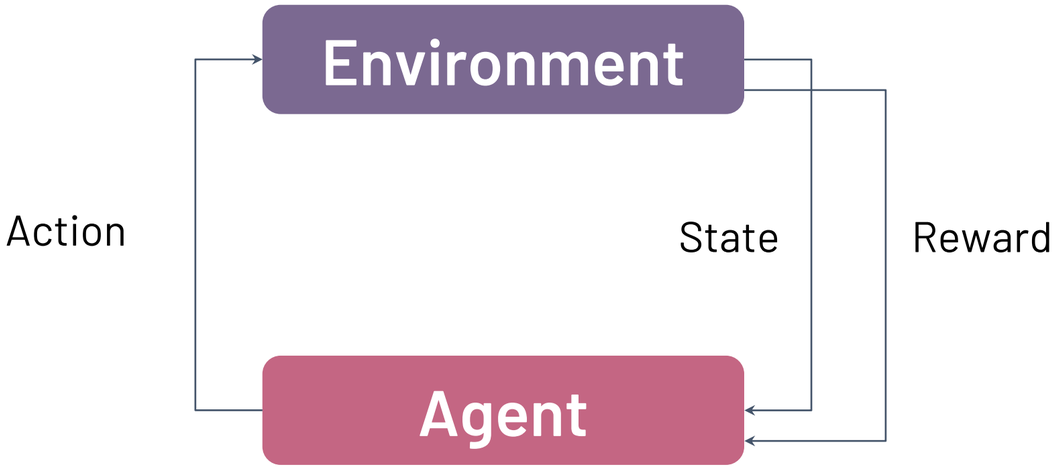
Reinforcement Learning (RL)

No labels: agent never told right or wrong

Agent interacts with environment (simulator or real world)

Typically can gather data, possibly at cost, by interacting with environment

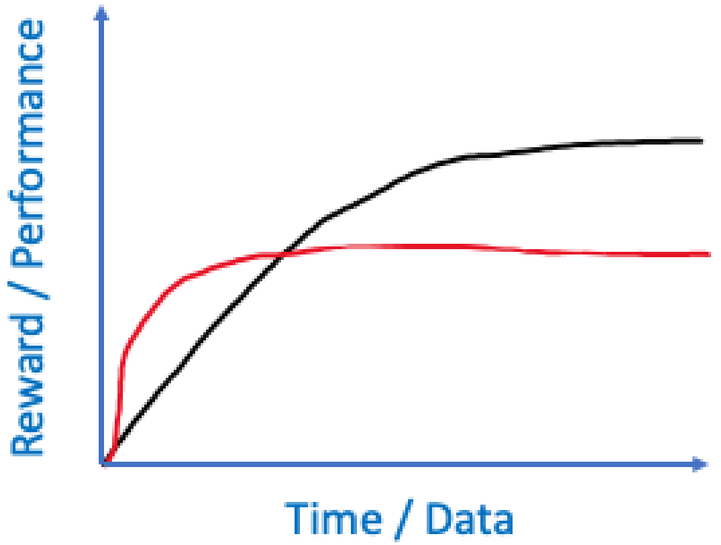
Learns via exploring vs. exploiting



RL Goals

Learn to maximize real-valued reward signal

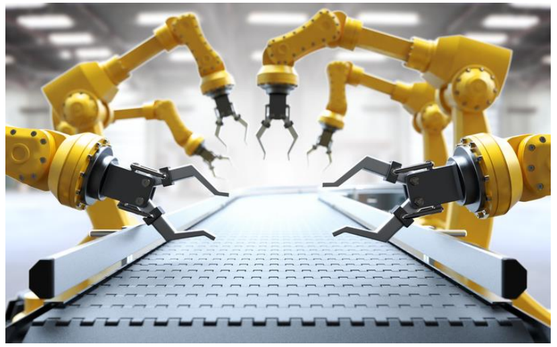
- With maximal final performance
- With little data
- Reducing human effort
- Discovering novel solutions
- Handling non-stationary environments



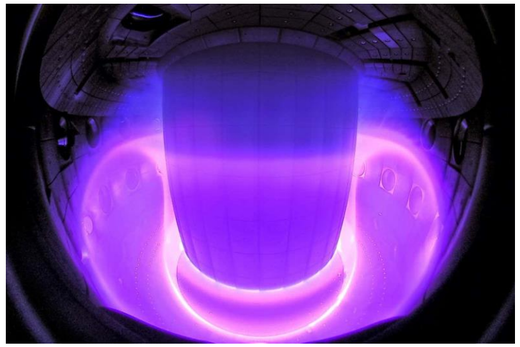
RL Applications

(Un)Supervised learning performs well for many real-world applications

Dota



Robotics



Fusion: Tokamak Plasmas

Stock Trading

AlphaGO

Data Center Cooling

https://www.youtube.com/watch?v=0Jw4HTWvGdY

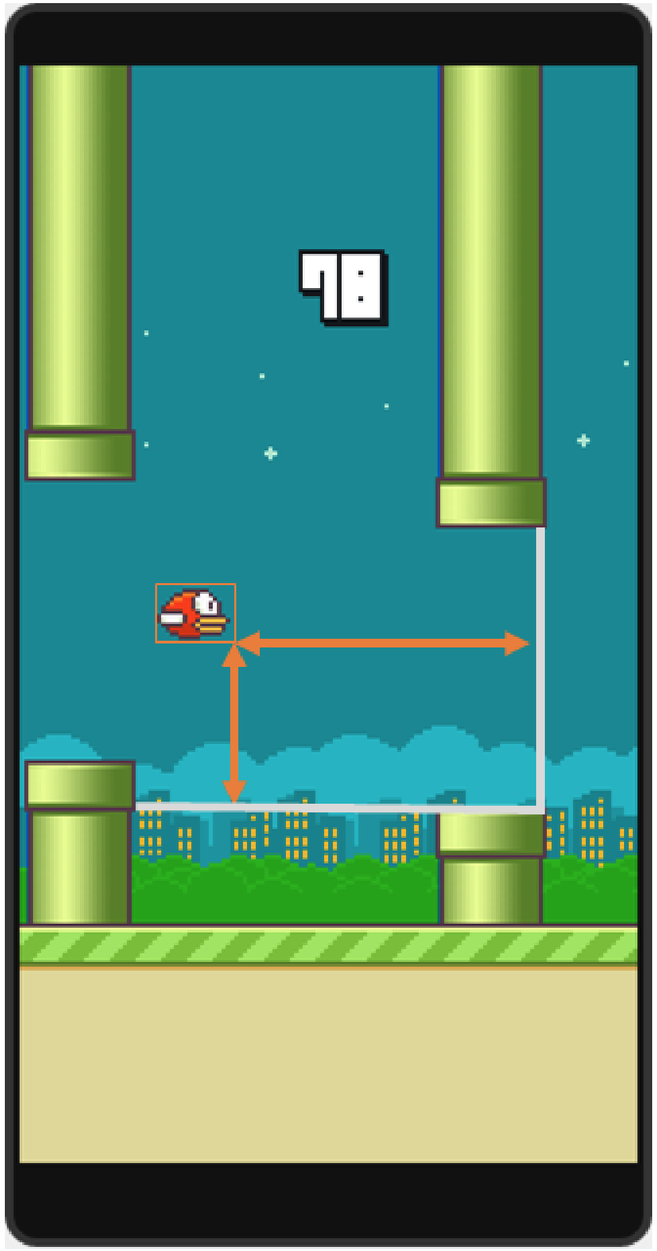
Example 1: Flappy Bird

Transition function: Controlled by game

Action?

Reward?

State representation?



http://sarvagyavaish.github.io/FlappyBirdRL/

Example 2: Aiden

https://www.borealisai.com/en/applying-ai/aiden/

Optimal Order Execution

State: Info about stock & market

Actions: Do nothing, buy/sell a little, buy/sell a lot

Rewards: Based on VWAP (Volume-weighted average price)

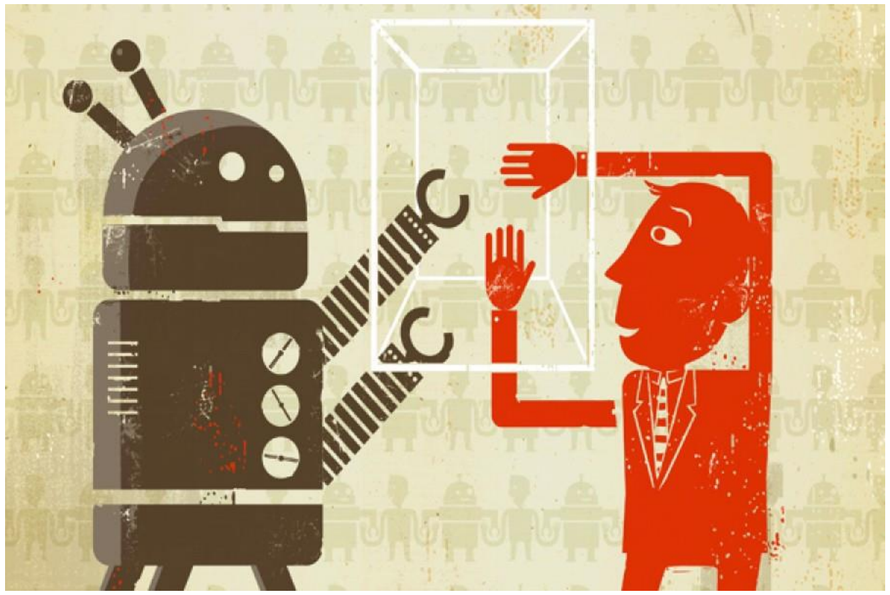
Transition function: Stock market (real or simulated data)

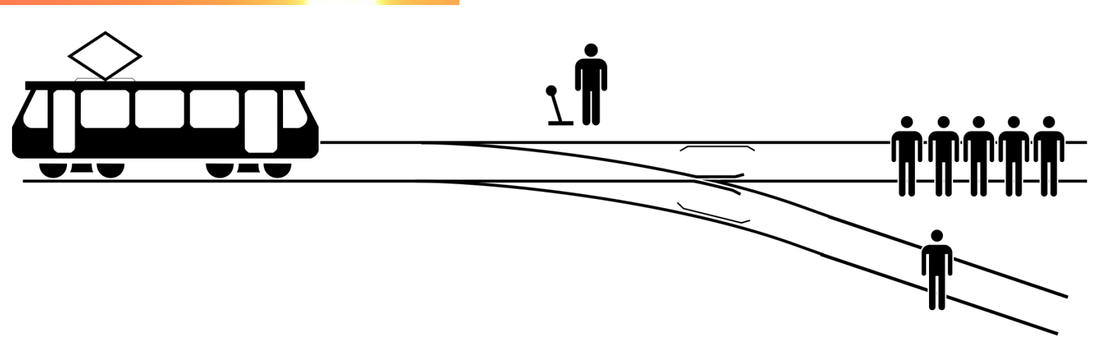
Ideas for Sequential Decision Processes?

Advantages of AI & Autonomy

Advantages of AI & Autonomy

Potential Disadvantages





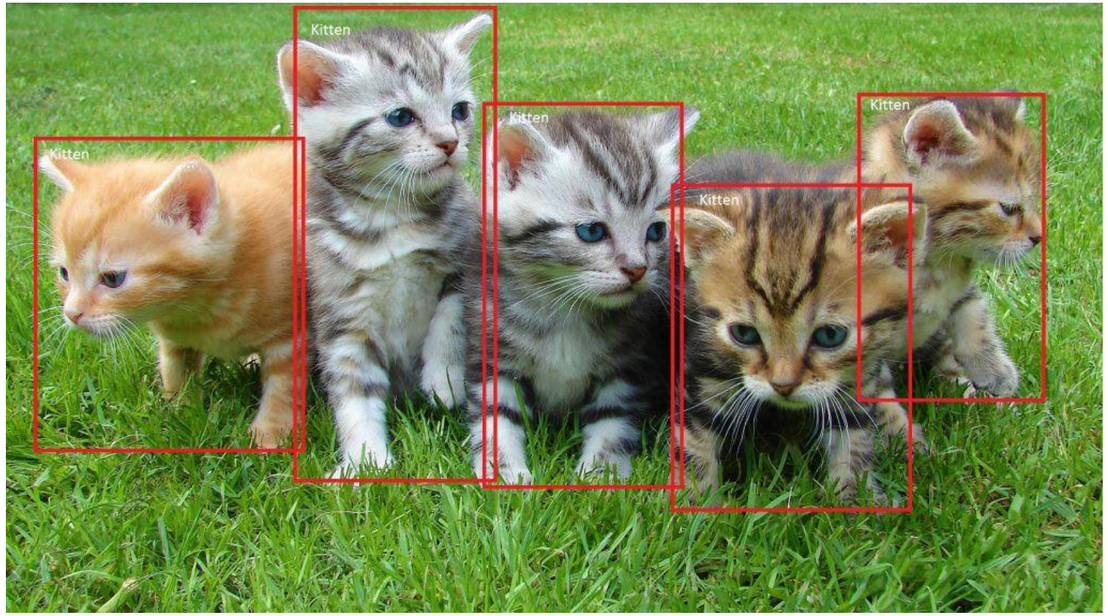
amazon mechanical turk™

Artificial Artificial Intelligence

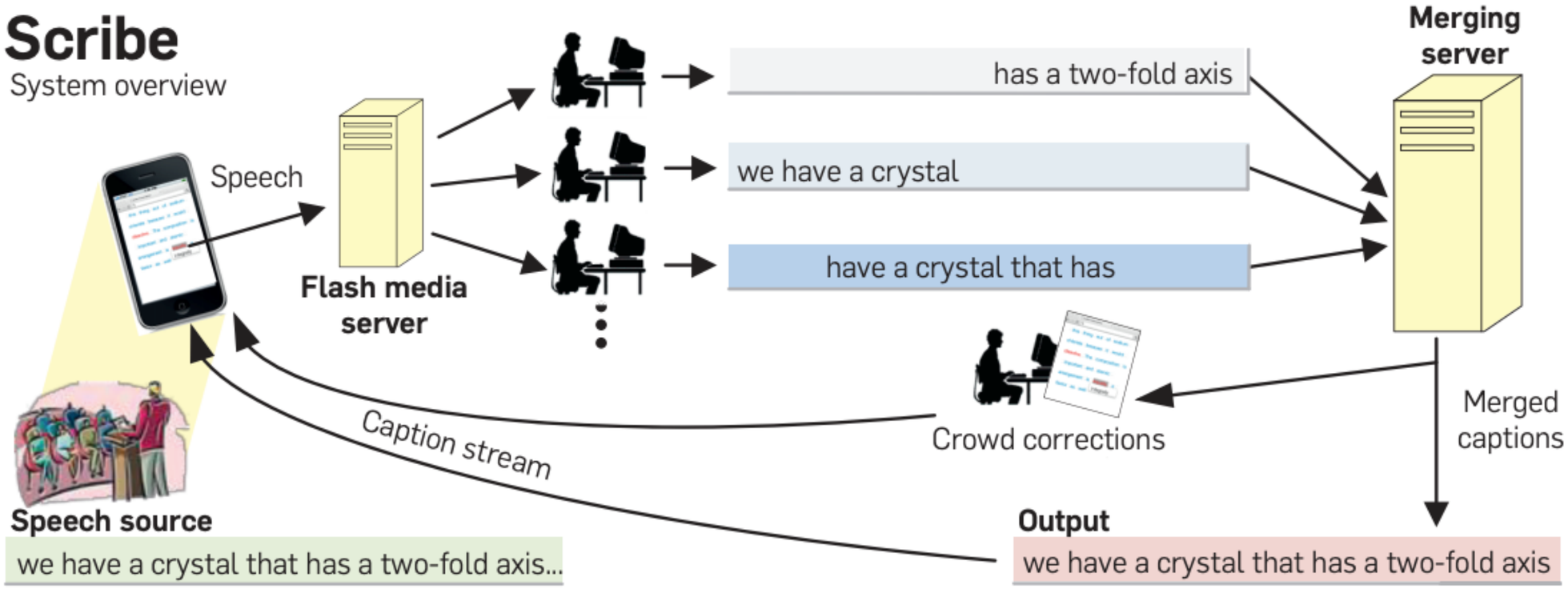
○ Cat
◉ Not Cat
○ Maybe/NotSure

◉ Cat
○ Not Cat
○ Maybe/NotSure

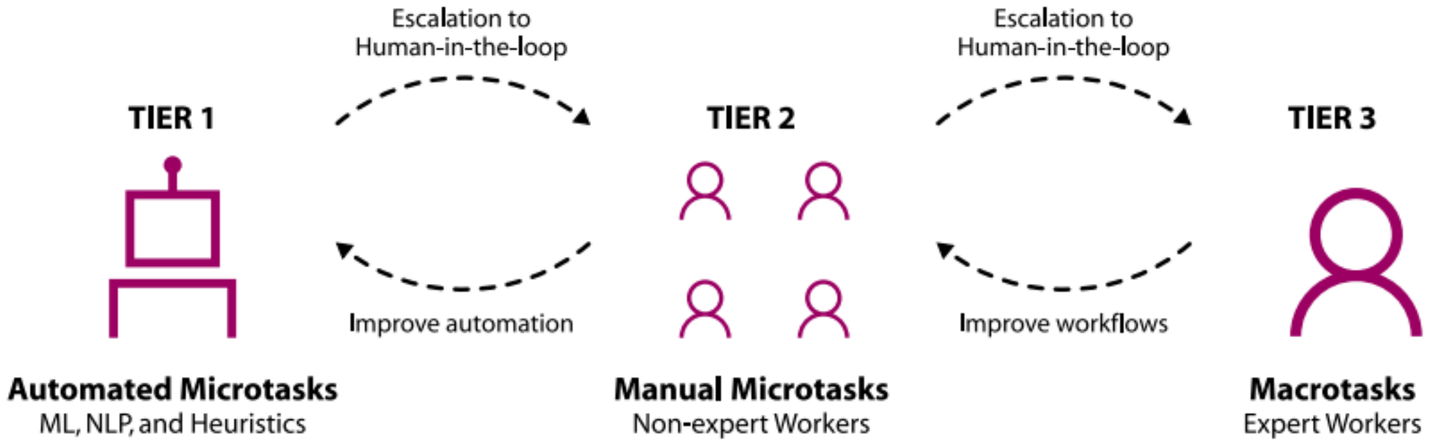
○ Cat
○ Not Cat
◉ Maybe/NotSure



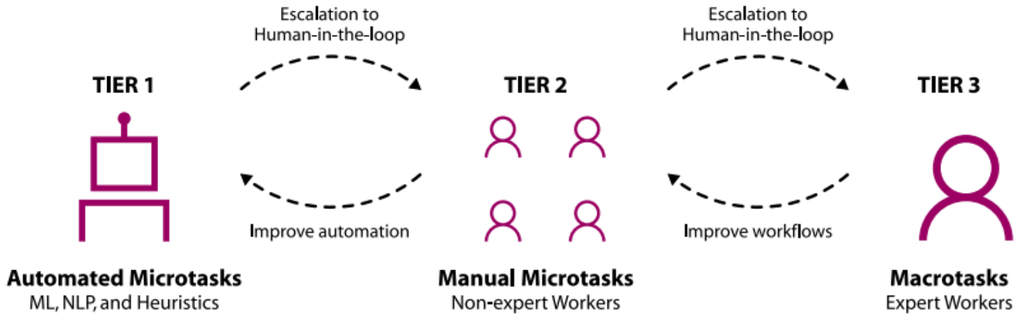
Legion & Scribe

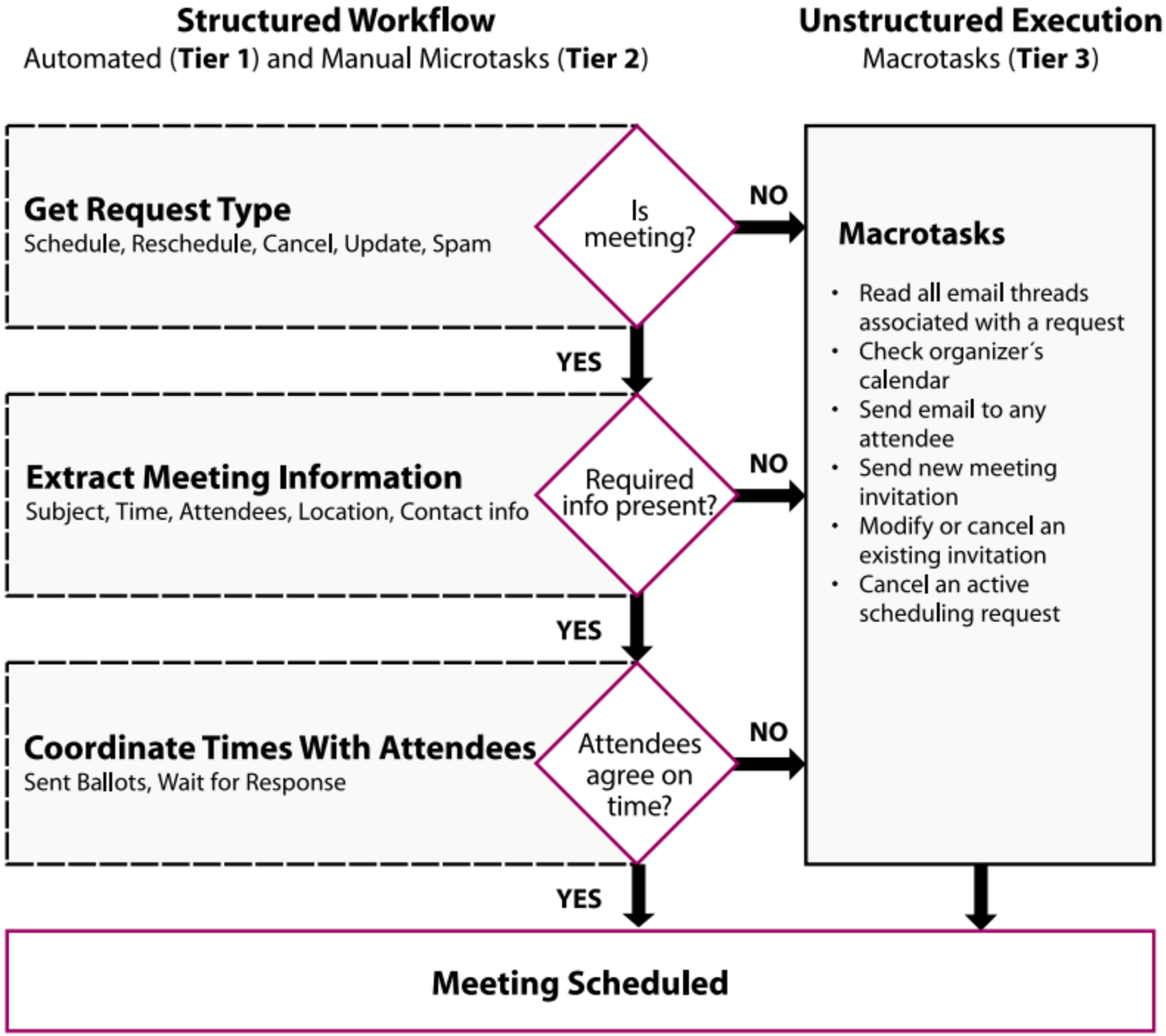


Calendar.help



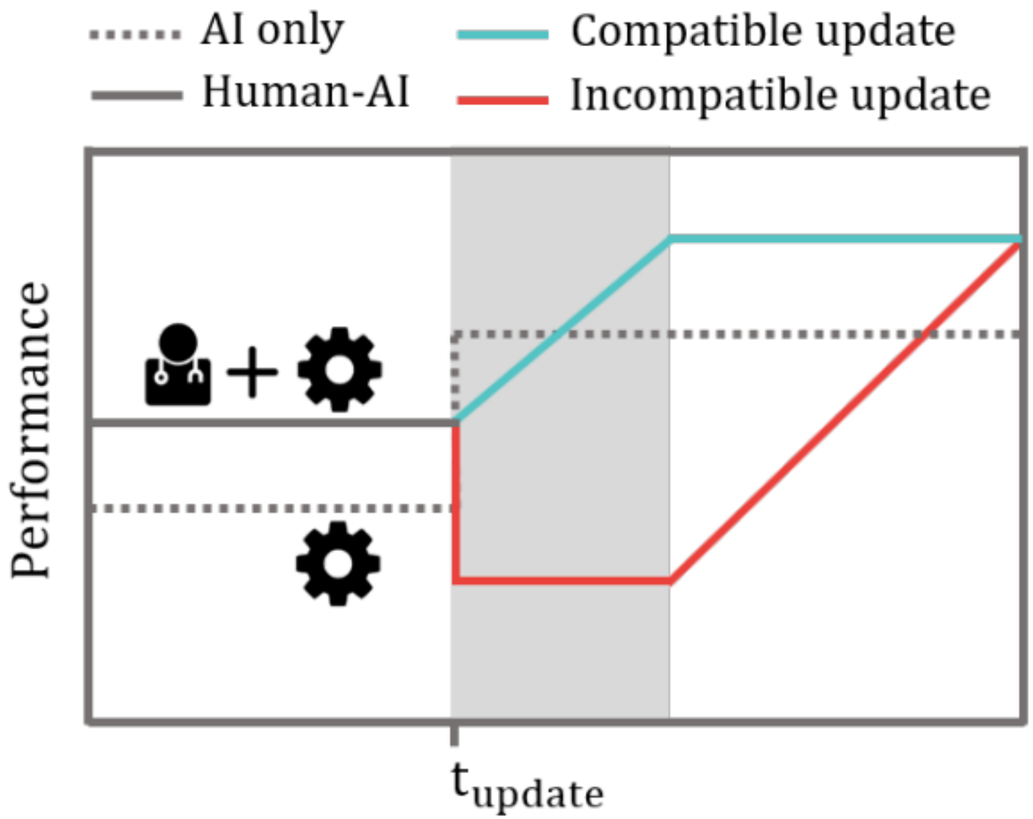
Calendar.help



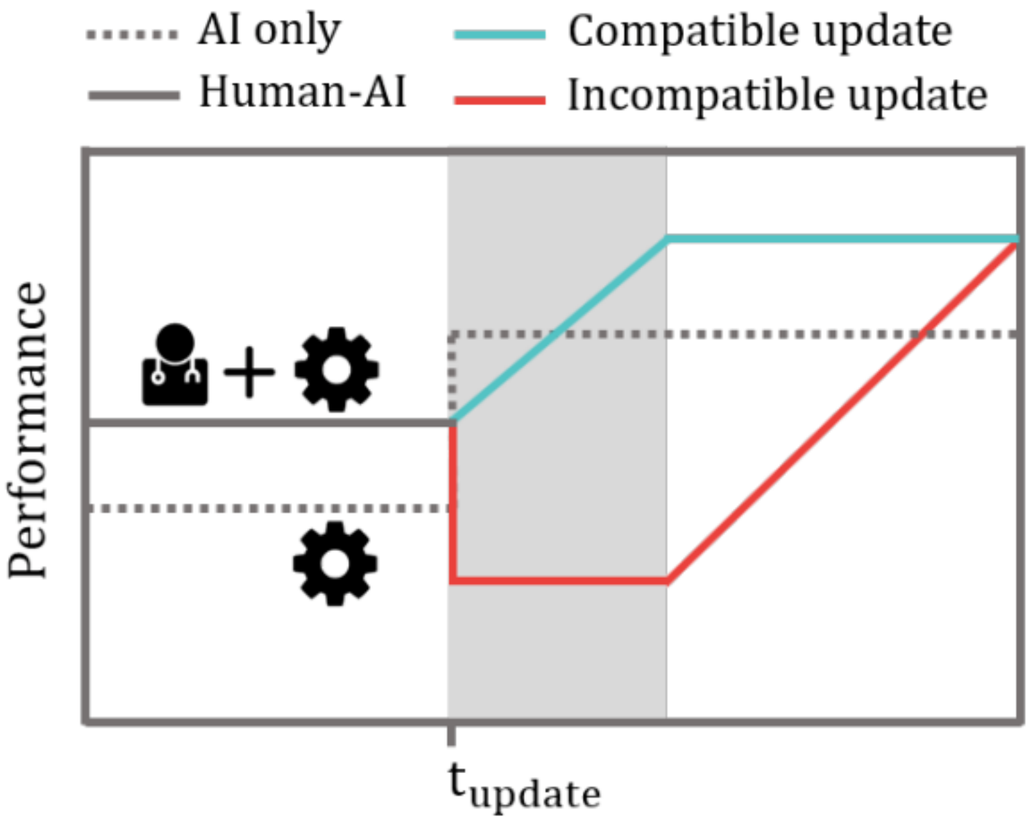


Robotics Teamwork

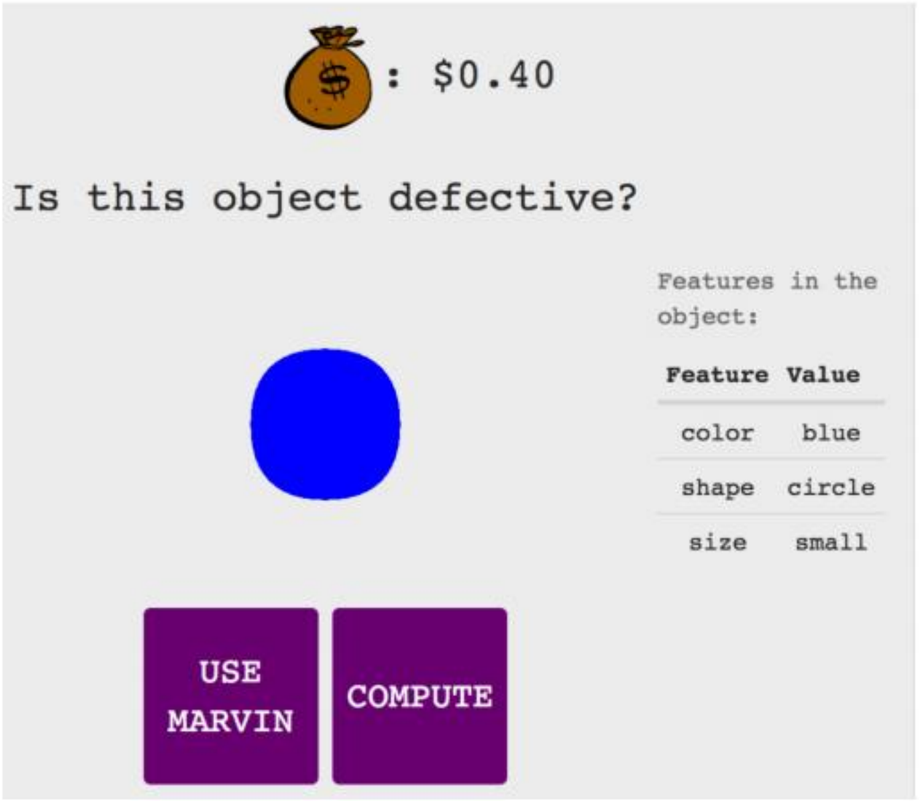
Human-AI Teams: Performance/Compatibility Tradeoff



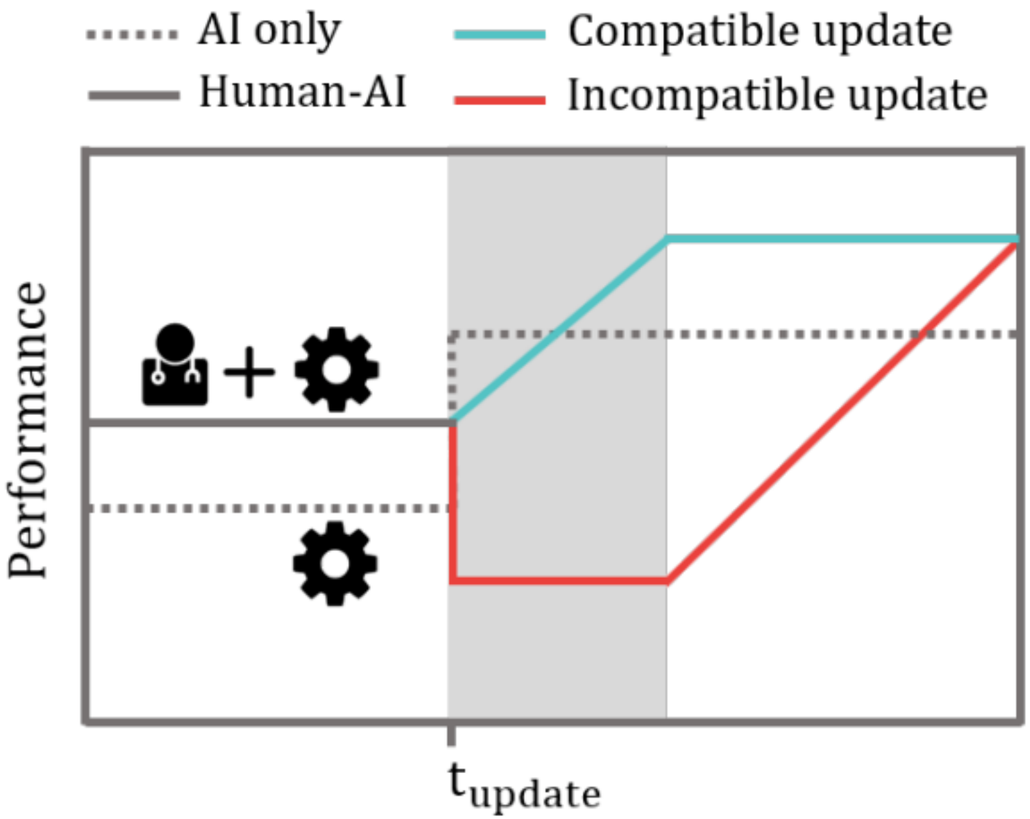
Human-AI Teams: Performance/Compatibility Tradeoff



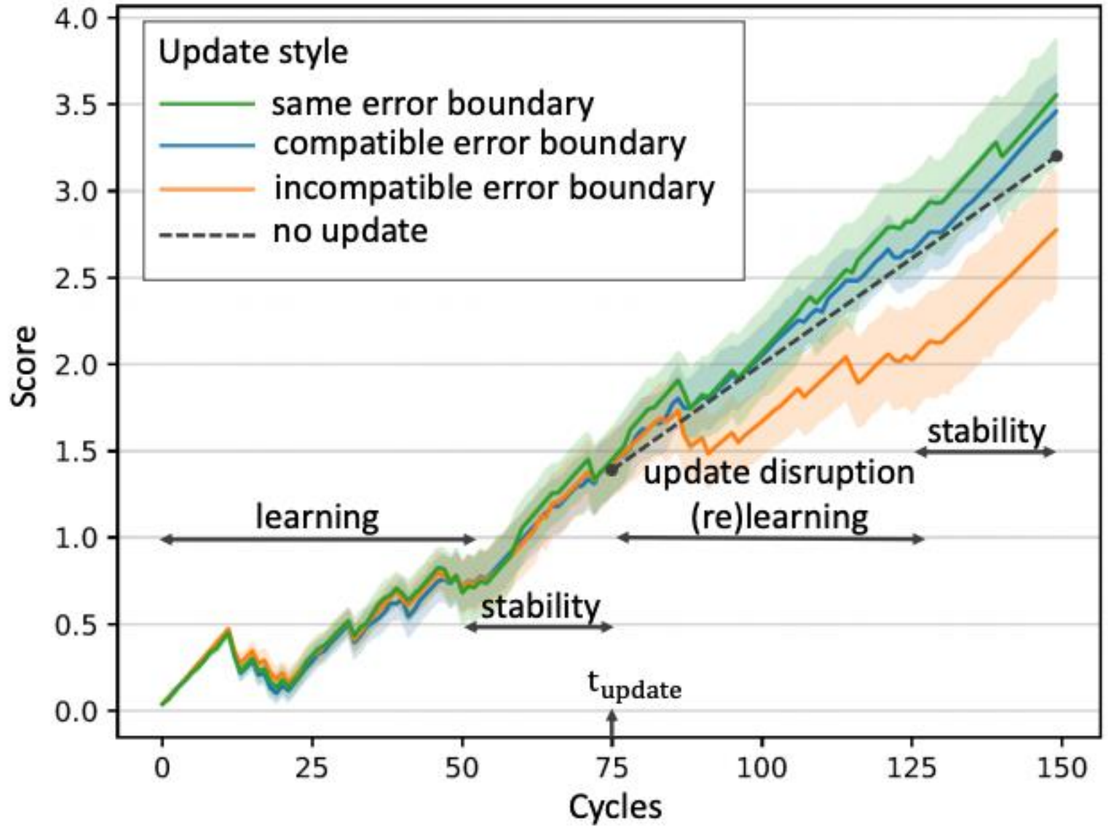
	Accept	Compute
AI right	$0.04	0
AI wrong	-$0.16	0



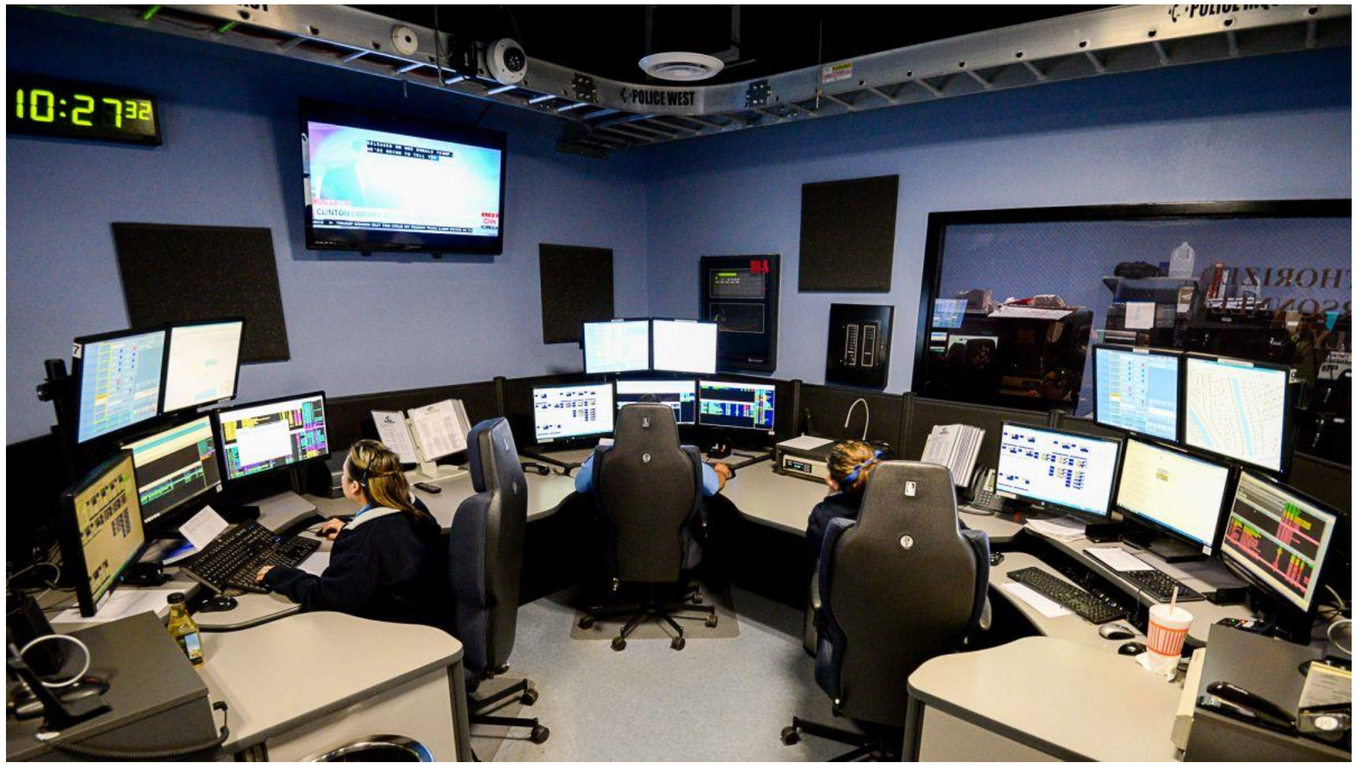
Human-AI Teams: Performance/Compatibility Tradeoff



	Accept	Compute
AI right	$0.04	0
AI wrong	-$0.16	0

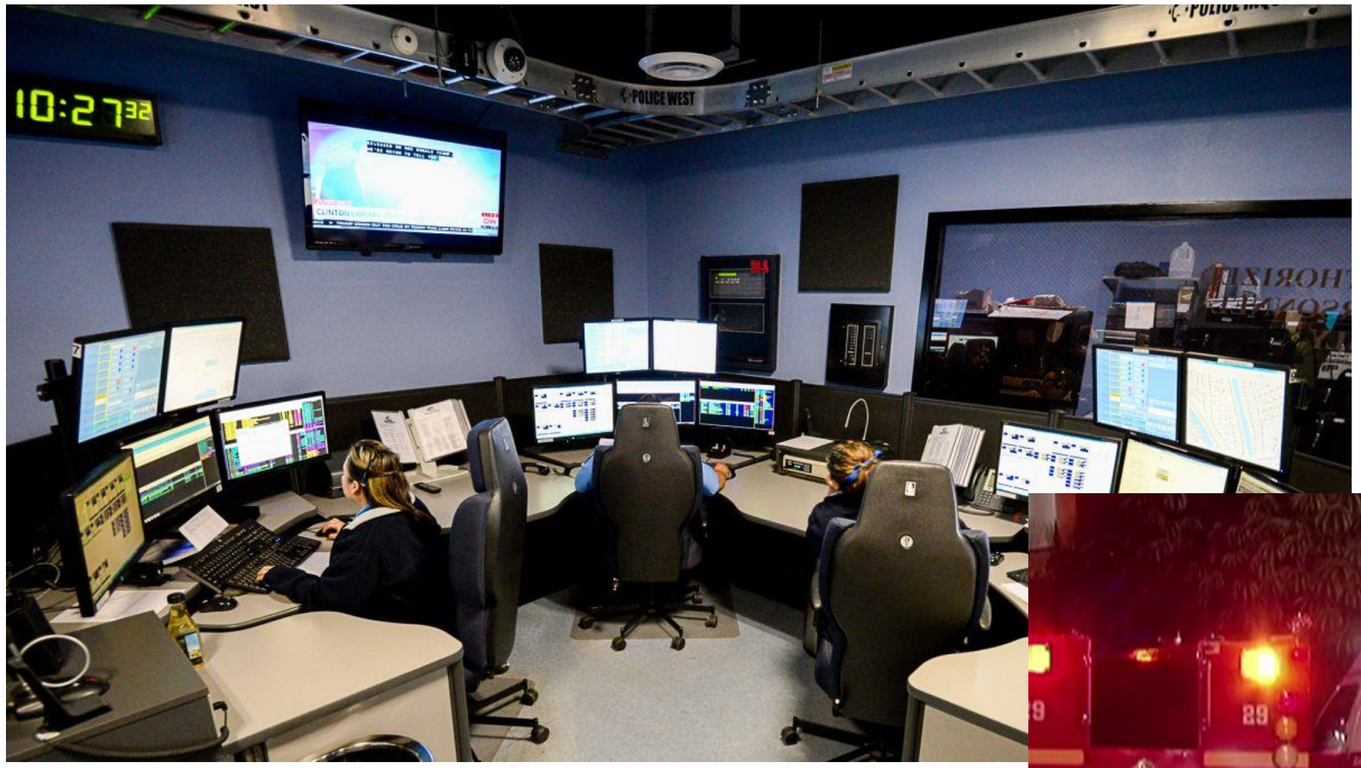


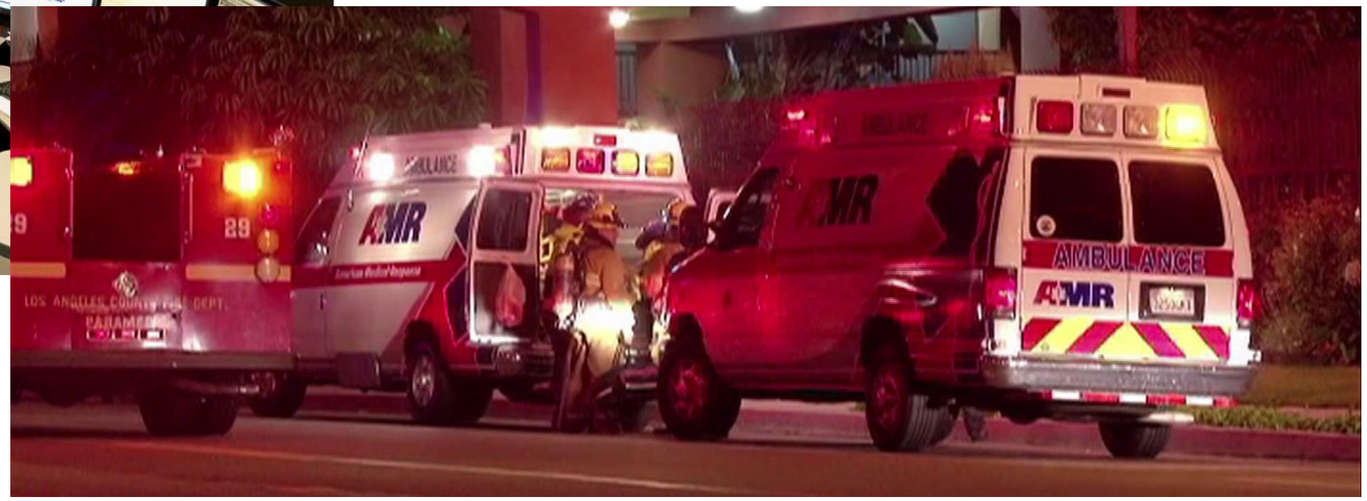
Emergency Vehicle Dispatching



Emergency Vehicle Dispatching

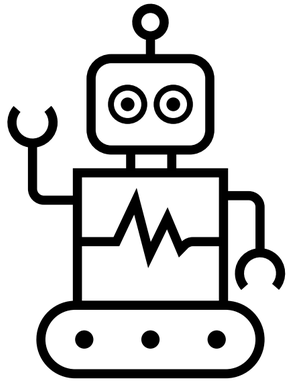
air
AI Redefined



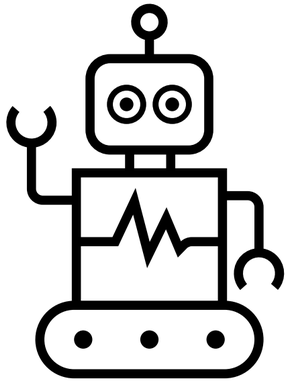


Question 1: What are the right RL problems?

- Sequential decision task
- High impact
- Robust to exploration

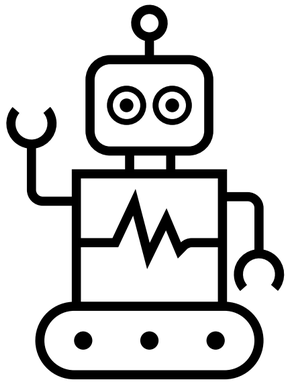


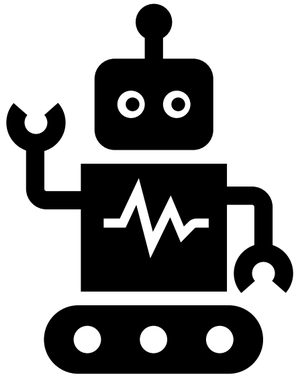
Question 2: Autonomous, Learning from, or Teaming with Humans?



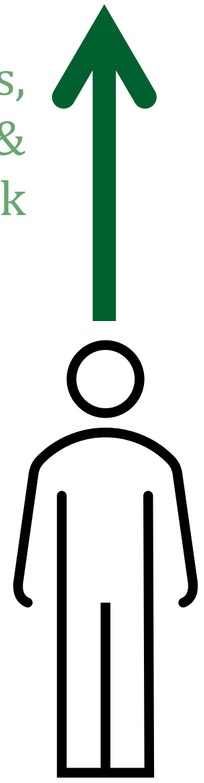
Question 2: Autonomous, Learning from, or Teaming with Humans?

Question 2: Autonomous, Learning from, or Teaming with Humans?

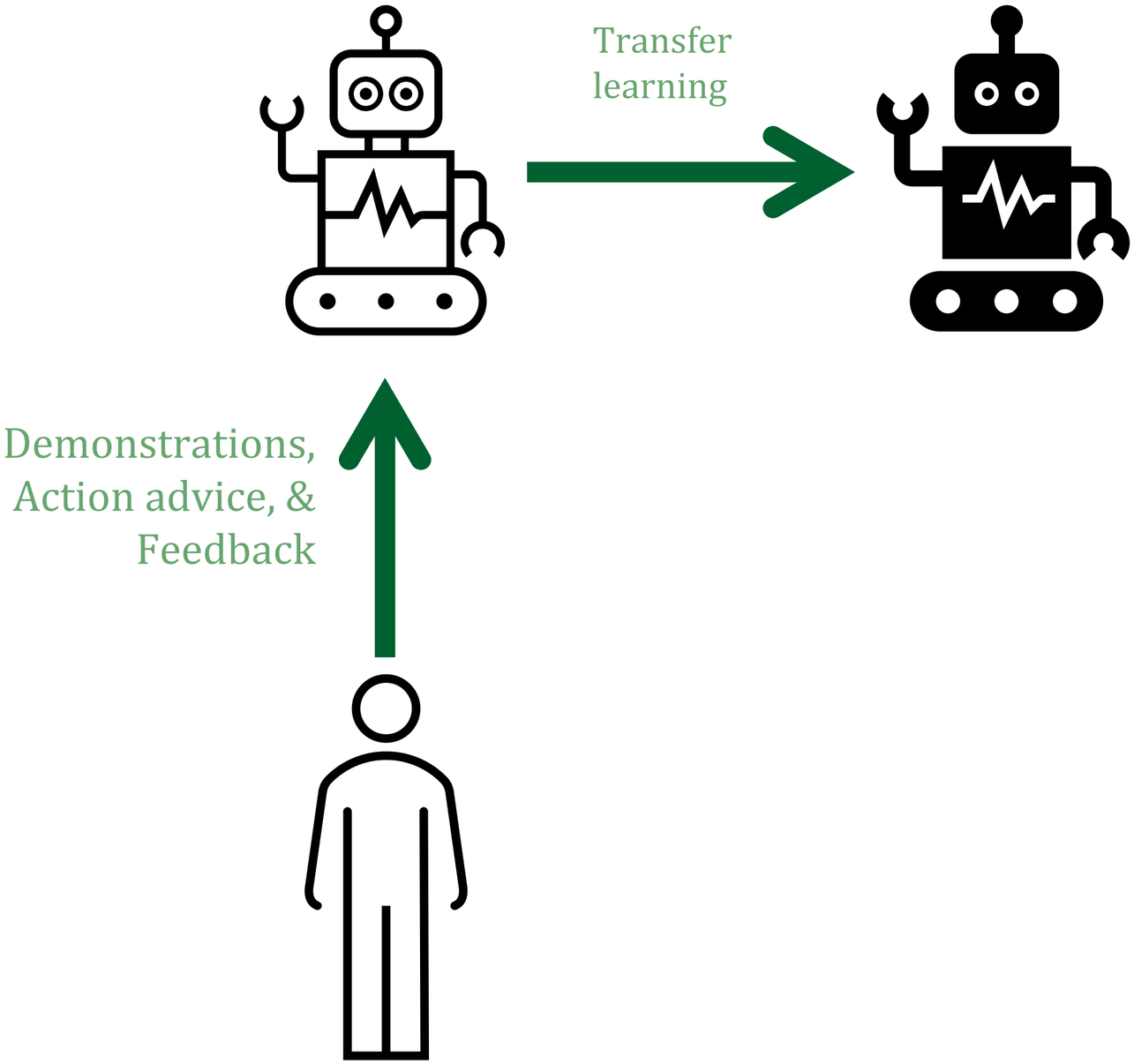




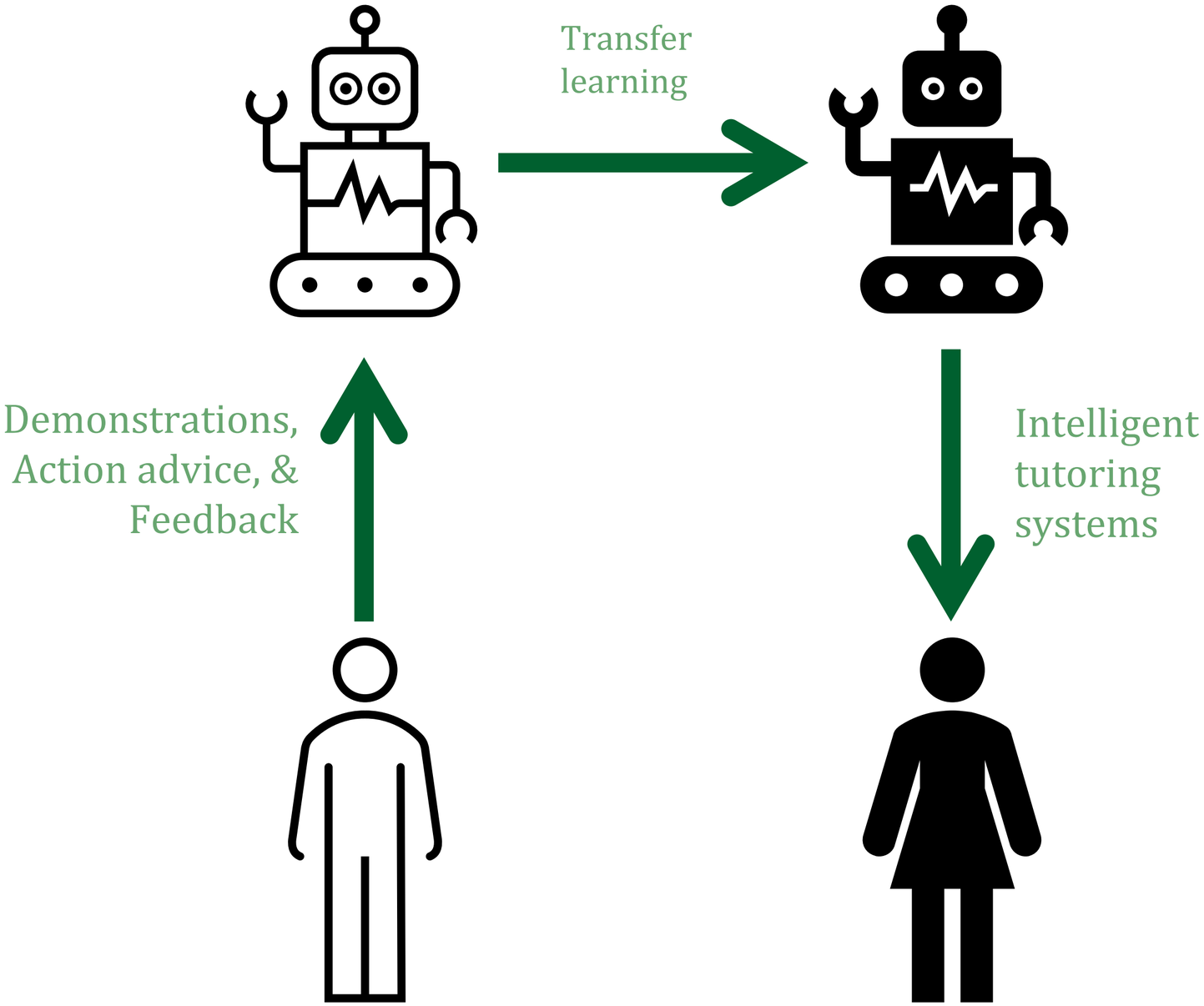
Demonstrations,
Action advice, &
Feedback



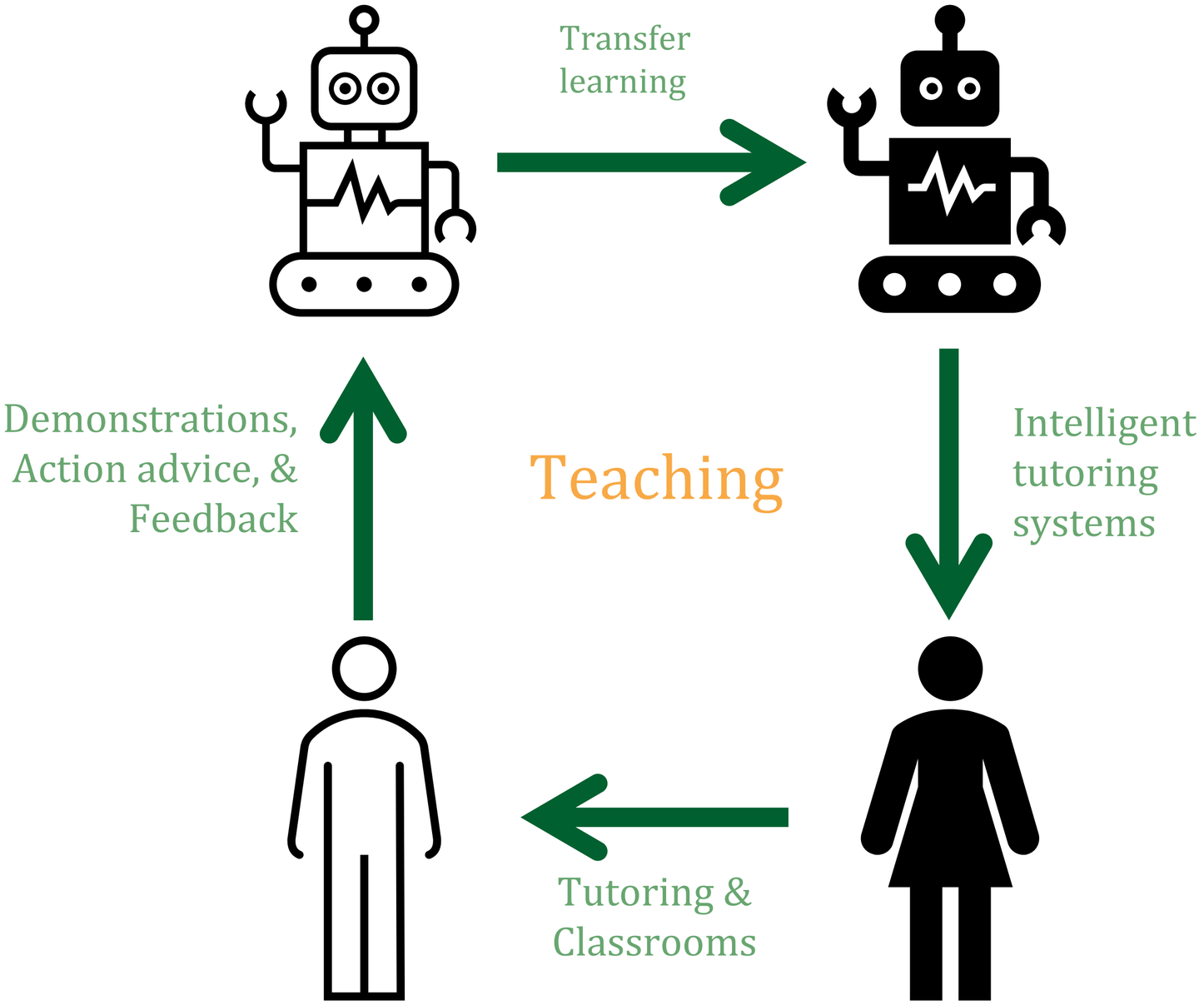
Question 2: Autonomous, Learning from, or Teaming with Humans?

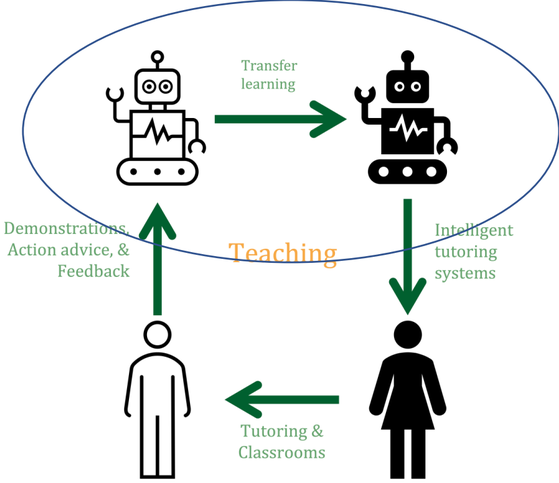


Question 2: Autonomous, Learning from, or Teaming with Humans?



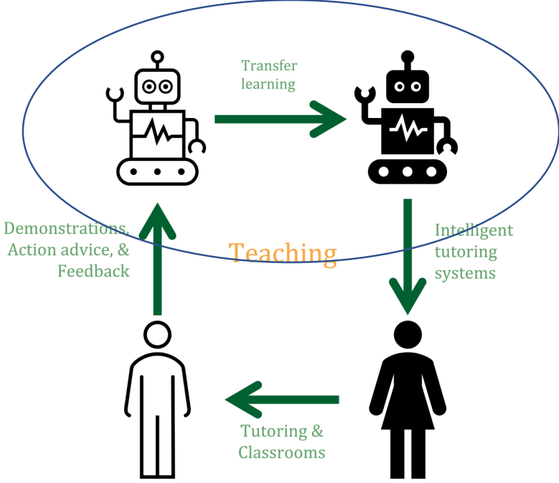
Question 2: Autonomous, Learning from, or Teaming with Humans?




Transfer
learning
Demonstrations,
Action advice, &
Feedback
Teaching
Intelligent
tutoring
systems
Tutoring &
Classrooms

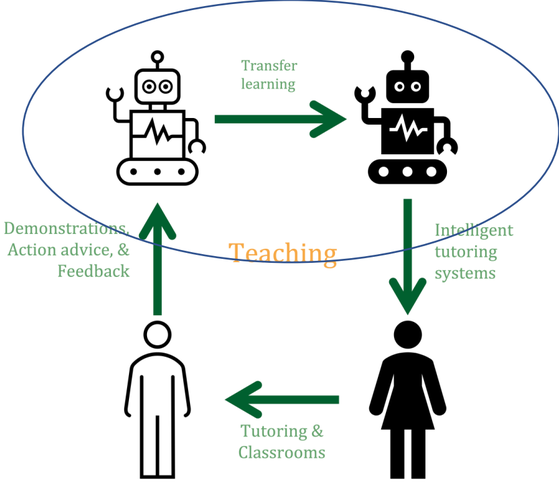
Agent Student, Agent Teacher

- Transfer Learning



Agent Student, Agent Teacher

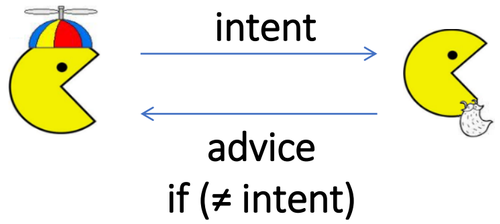
- Transfer Learning
- HAT ← Teacher Decides



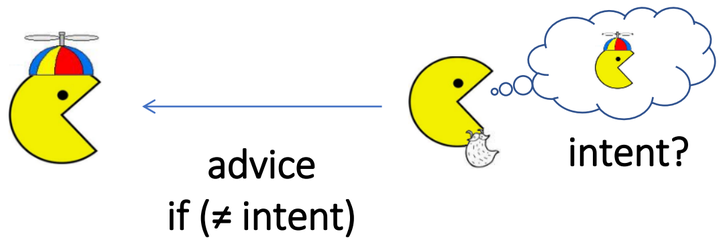
Early advising

Importance advising

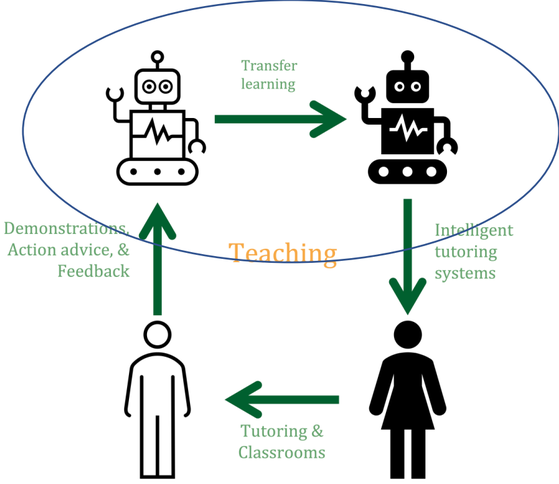
Mistake correcting



Predictive advising



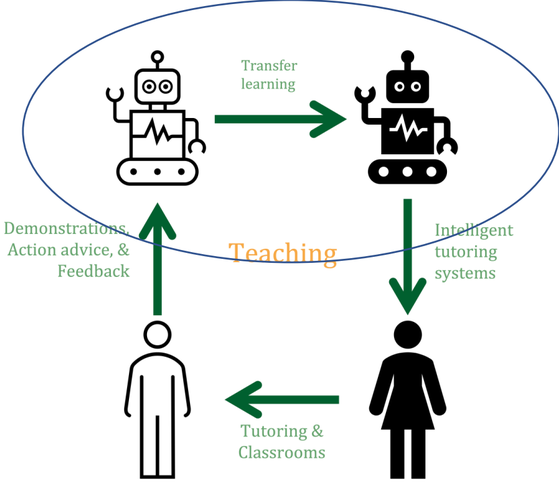
Agent Student, Agent Teacher



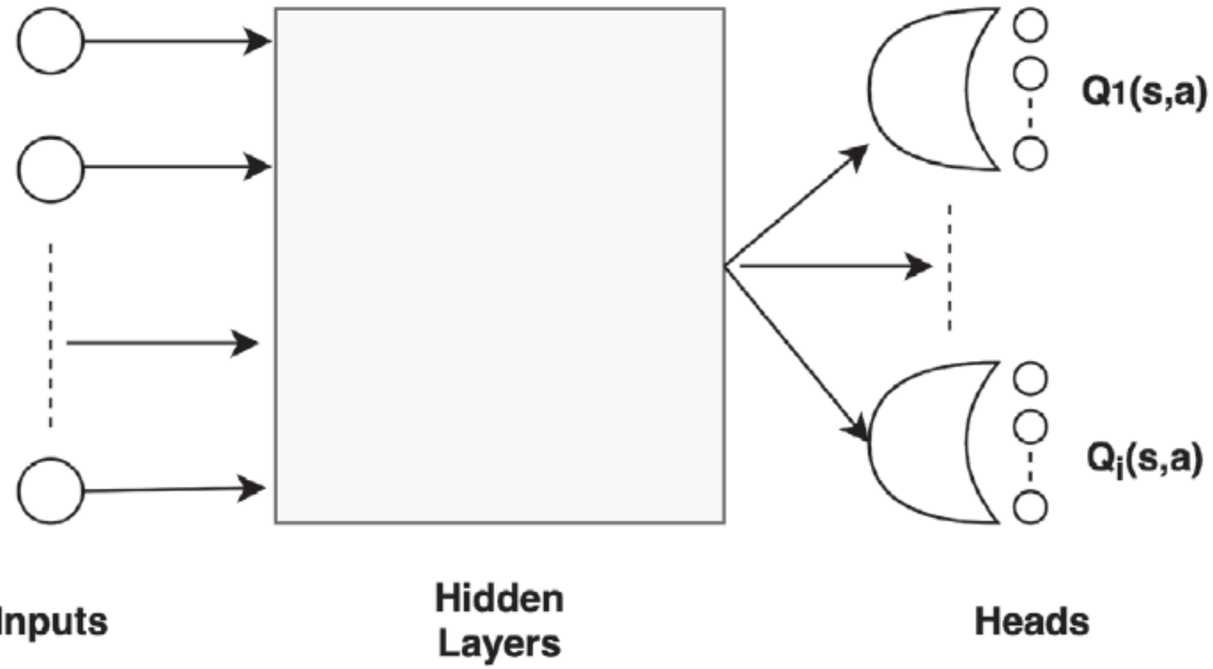
- Transfer Learning
- HAT ← Teacher Decides
- LeCTR (Omidshafiei+, 2018) ← Teacher Learns

Advising Reward Name	Description
JVG: Joint Value Gain	Task-level value $V(s;\boldsymbol{\theta})$ improvement after learning
QTR: Q-Teaching Reward	Teacher's estimate of best vs. intended student action
LG: Loss Gain	Student's task-level loss $\mathcal{L}(\theta^i)$ reduction
LGG: Loss Gradient Gain	Student's task-level policy gradient magnitude
TDG: TD Gain	Student's temporal difference (TD) error δ^i reduction
VEG: Value Estimation Gain	Student's value estimate $\hat{V}(\theta^i)$ gain above threshold τ

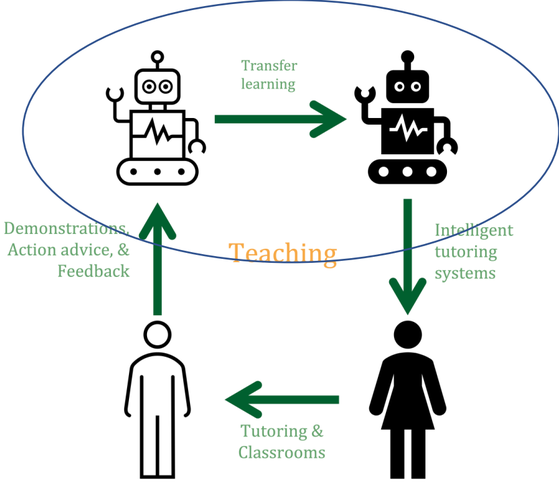
Agent Student, Agent Teacher



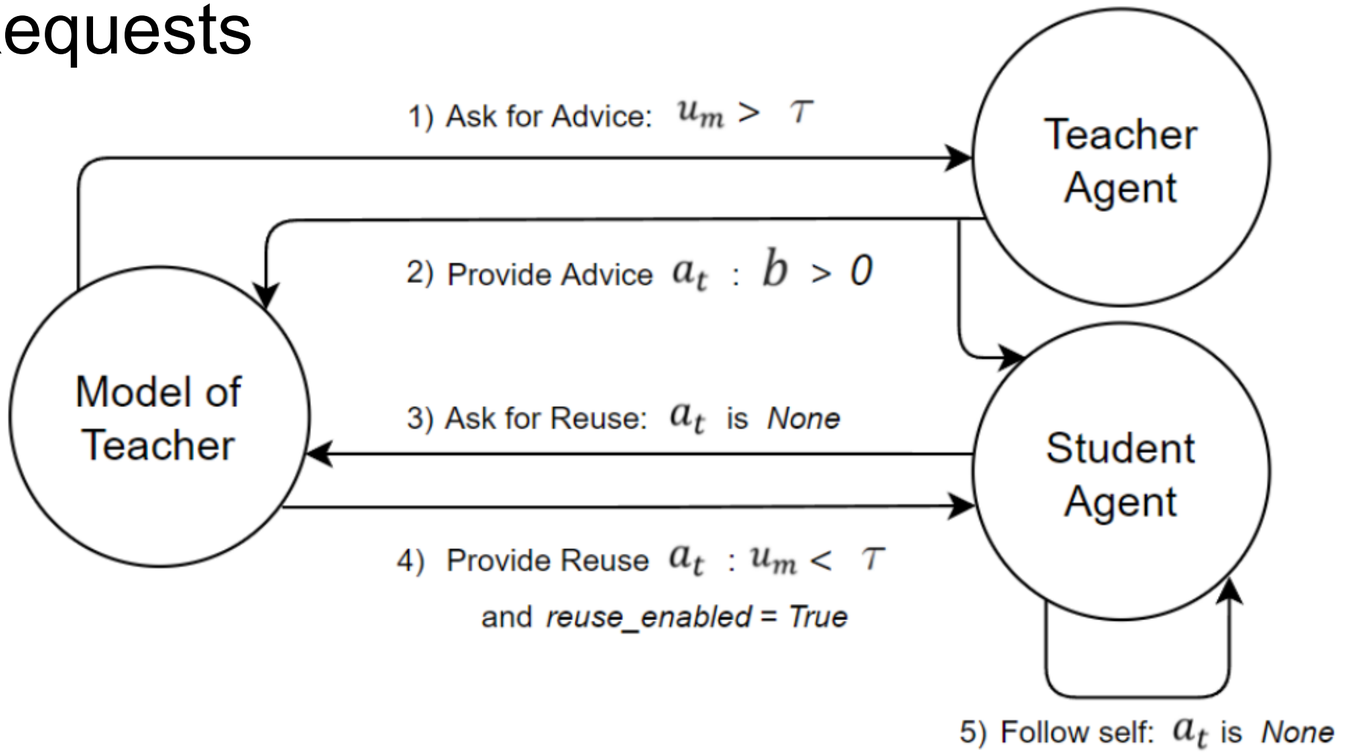
- Transfer Learning
- HAT ← Teacher Decides
- LeCTR (Omidshafiei+, 2018) ← Teacher Learns
- RCMP → Student Requests

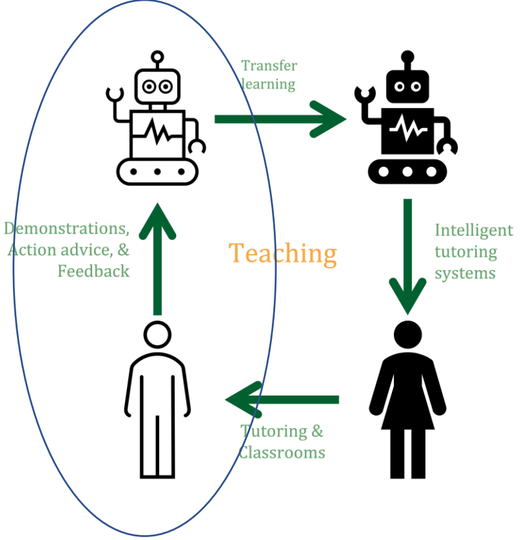


Agent Student, Agent Teacher

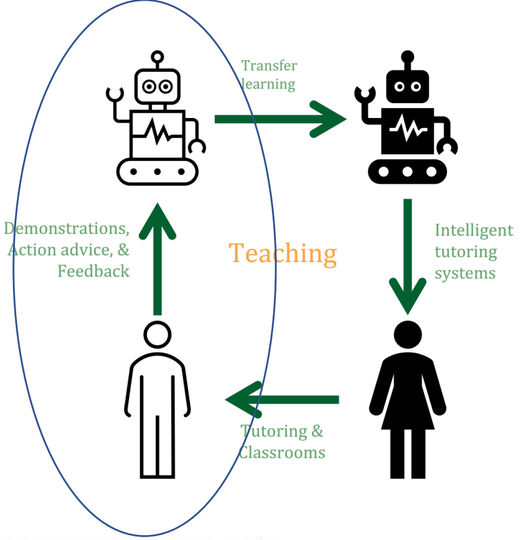


- Transfer Learning
- HAT ← Teacher Decides
- LeCTR (Omidshafiei+, 2018) ← Teacher Learns
- RCMP → Student Requests
- AIR (İlhan+, 2021) → Student Requests




Transfer
learning
Demonstrations,
Action advice, &
Feedback
Teaching
Intelligent
tutoring
systems
Tutoring &
Classrooms

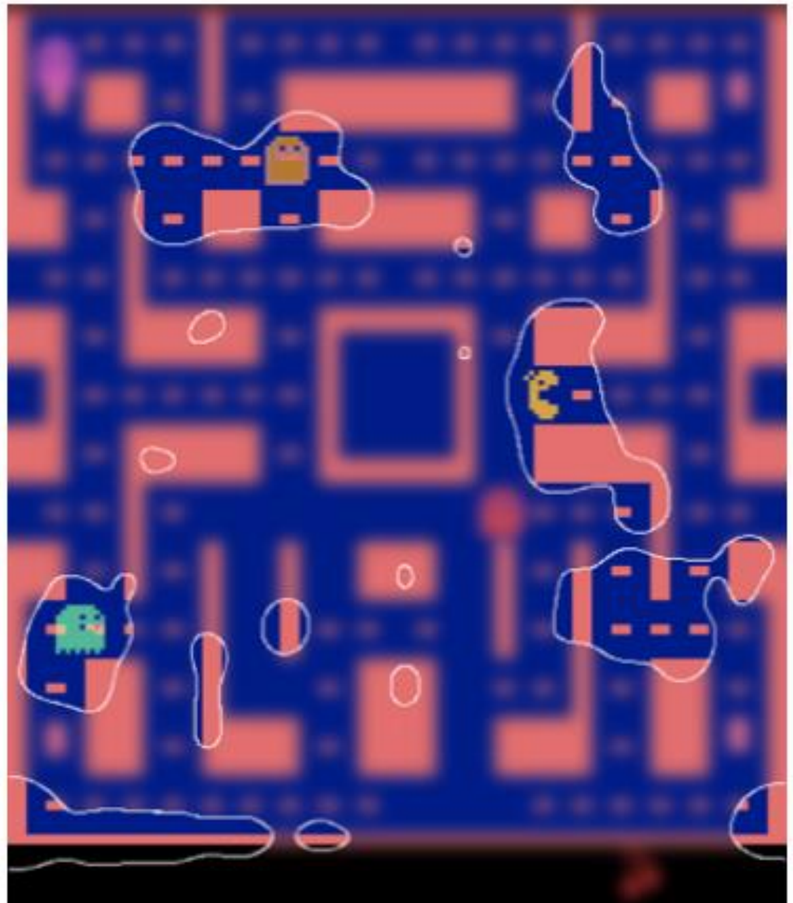
Agent Student, Human Teacher



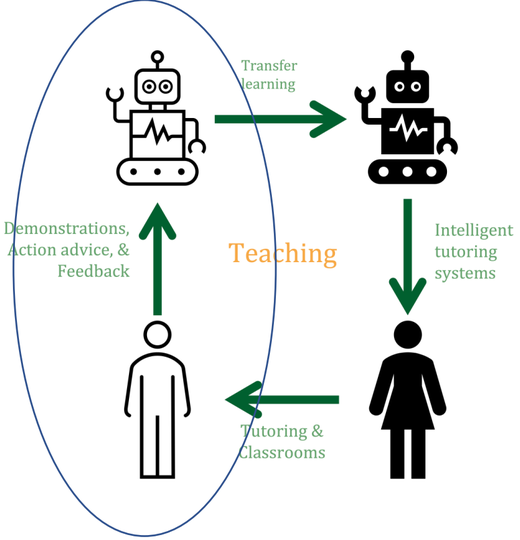
Agent can also provide Explanations

- “Why did you do A?”
- “Why didn’t you do B?”
- “What would happen if you did C?”

(Davis Pierson+, 2022)



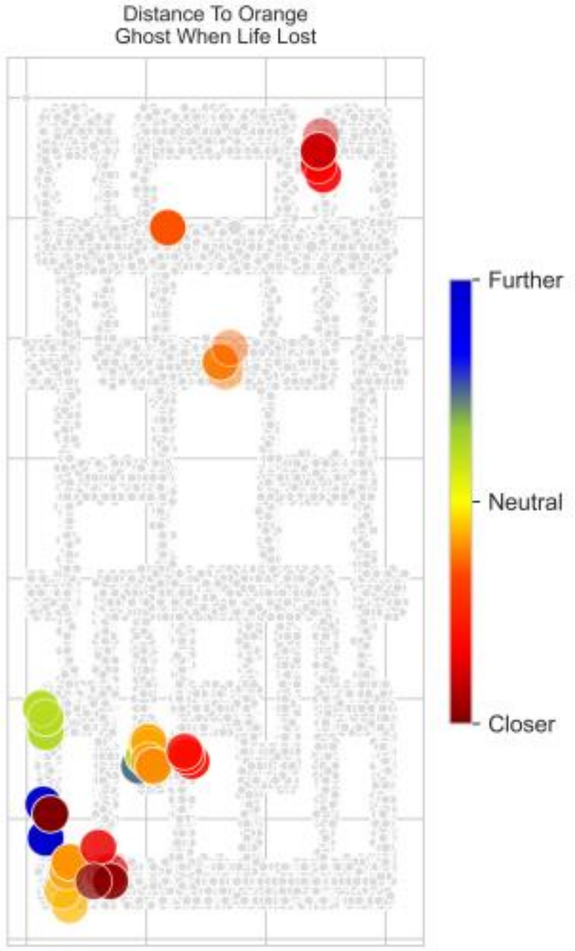
Agent Student, Human Teacher



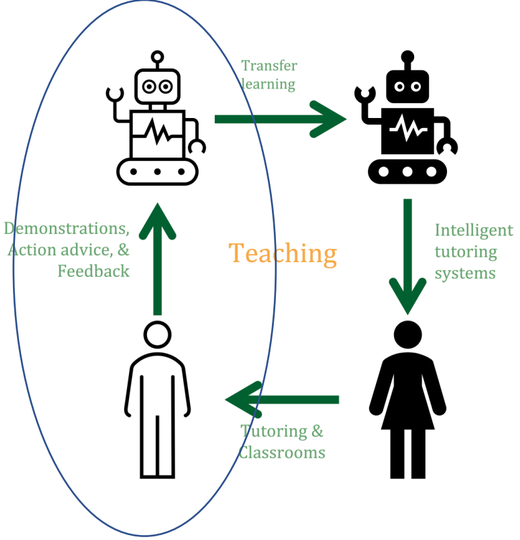
Agent can also provide Explanations

- “Why did you do A?”
- “Why didn’t you do B?”
- “What would happen if you did C?”

(Davis Pierson+, 2022)



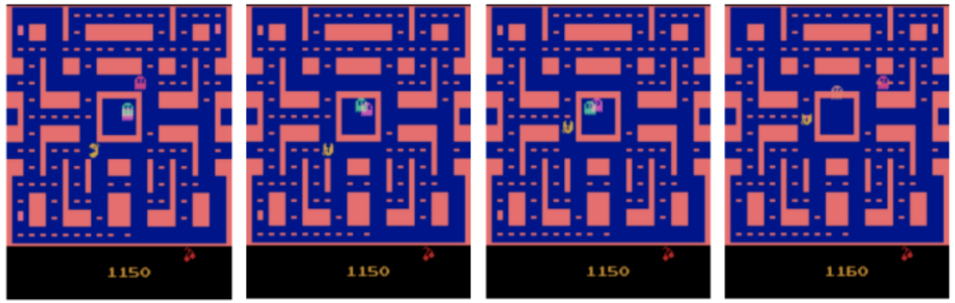
Agent Student, Human Teacher



Agent can also provide Explanations

- “Why did you do A?”
- “Why didn’t you do B?”
- “What would happen if you did C?”

(Davis Pierson+, 2022)



question

3/4 If you were in a competition where your agent needed to STAY ALIVE THE LONGEST, which AI agent would you select to play on your behalf?

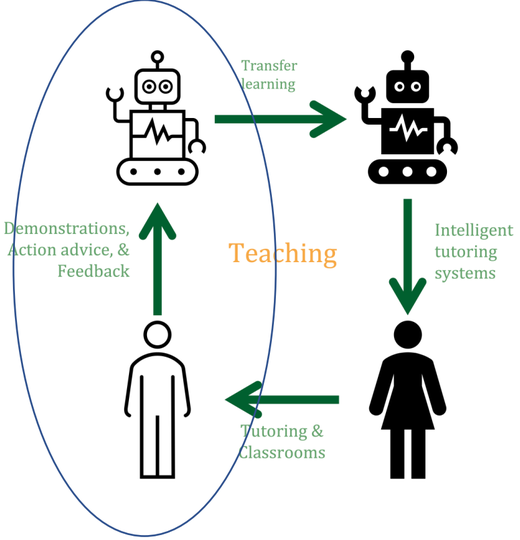
Agent Thompson | Agent Jackson

question

1/13 What will Agent Thompson do next in the situation described in the 4 images above?

Go Up | Go Down | Go Left | Go Right | I Do Not Know

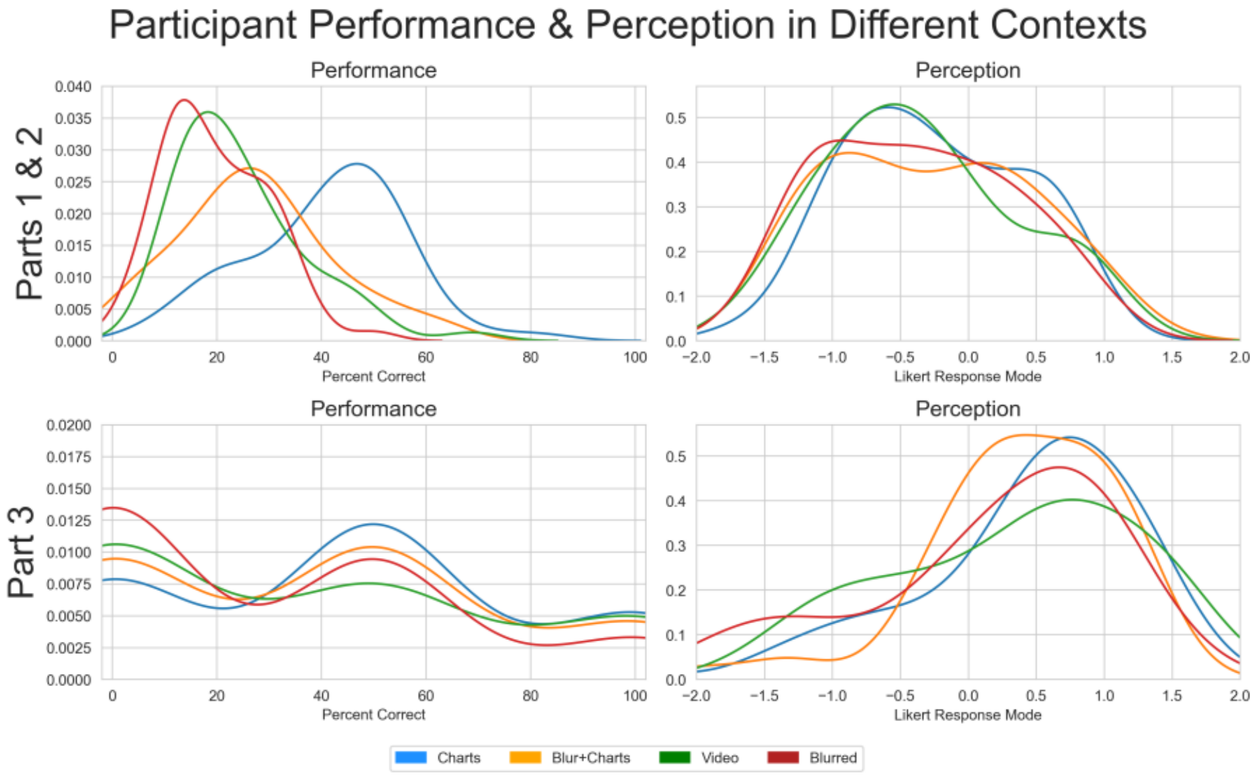
Agent Student, Human Teacher

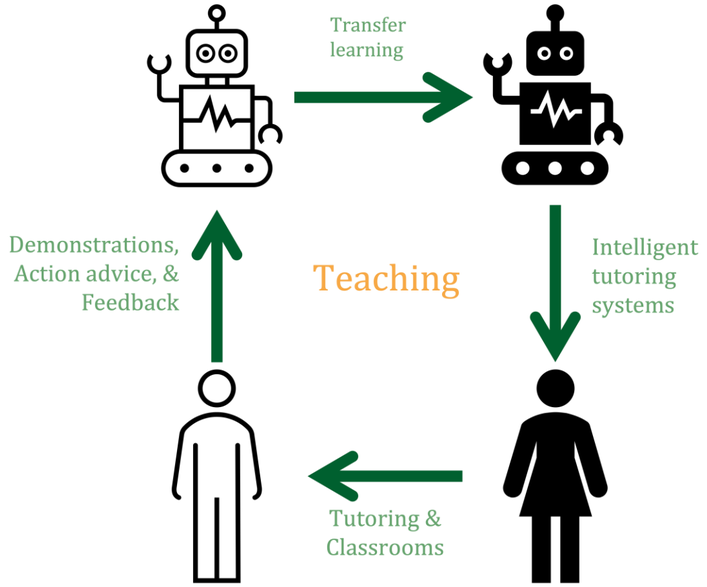


Agent can also provide Explanations

- “Why did you do A?”
- “Why didn’t you do B?”
- “What would happen if you did C?”

(Davis Pierson+, 2022)




Transfer
learning
Demonstrations,
Action advice, &
Feedback
Teaching
Intelligent
tutoring
systems
Tutoring &
Classrooms

Other open questions

Human-agent teaming

Multi-agent systems

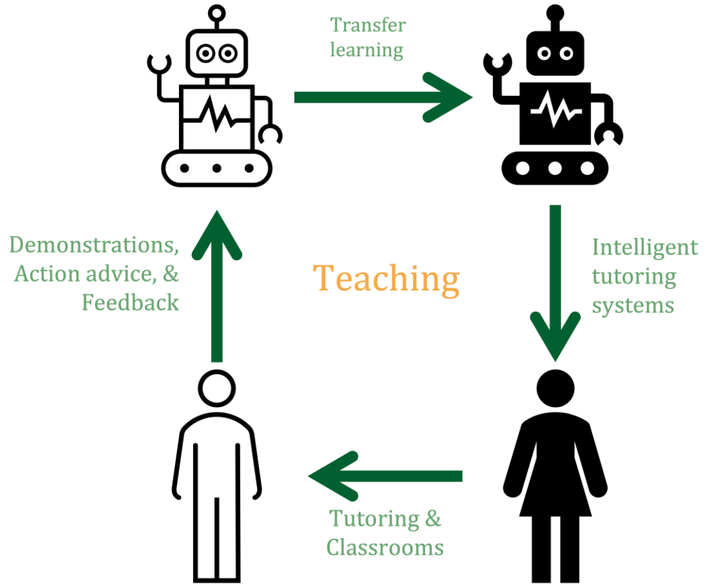
Multi-agent, multi-human systems

Explainability

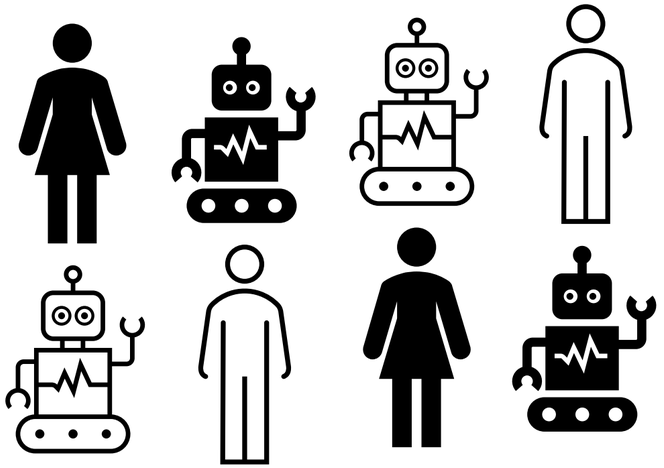
Accountability

Safety

My claim: Bar is low!



Next Steps



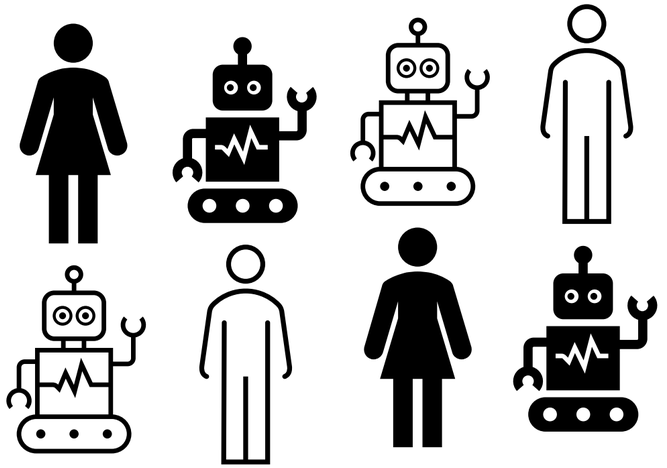
Sequential decision tasks?

Good task for machine intelligence?

Bootstrap off human knowledge?

Human-in-the-loop system?

Multi-agent, Multi-human Teaming



cogment™

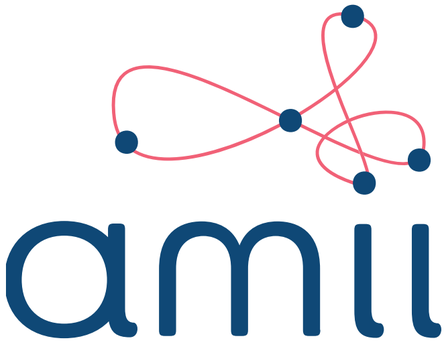
First released in 2019. Version 1.0 released in June 2021. Continuous updates released since, including Cogment Verse, a series of actors and environments to make Cogment faster and easier to use.

The **first platform** to allow the design, training, and deployment of complex **intelligence ecosystems**, mixing **humans and artificial agents** of various kinds

It orchestrates heterogeneous ML & non-ML agents with real-time human interaction.

air
AI Redefined

The Intelligent
Robot Learning
Laboratory



Thank you!

http://irll.ca

Matthew E. Taylor