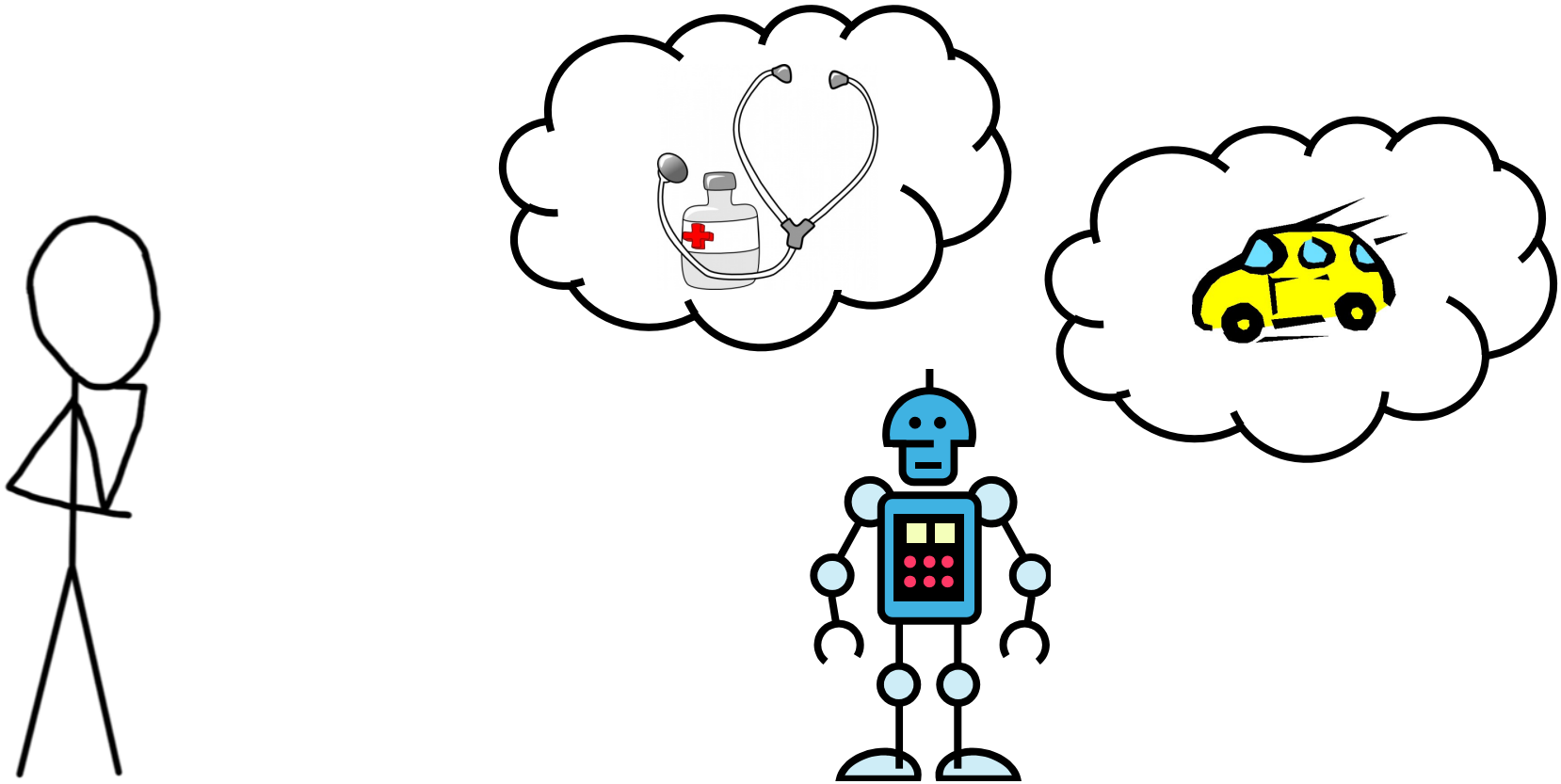
Toward Robust Summarization of Agent Policies

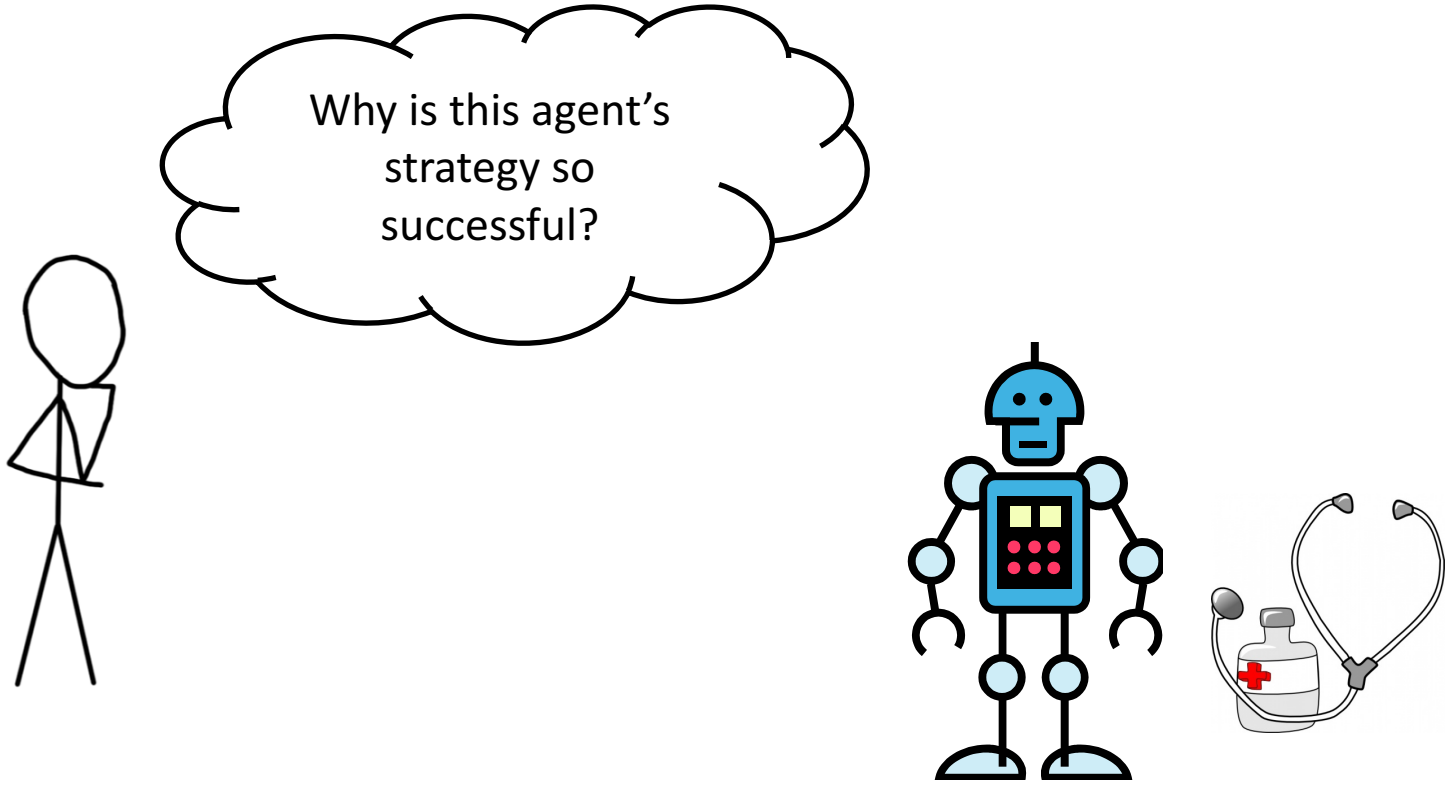
Isaac Lage[1], Daphna Lifschitz[2],
Finale Doshi-Velez[1] & Ofra Amir[2]

[1]Harvard University
[2]Technion – Israel Institute of Technology

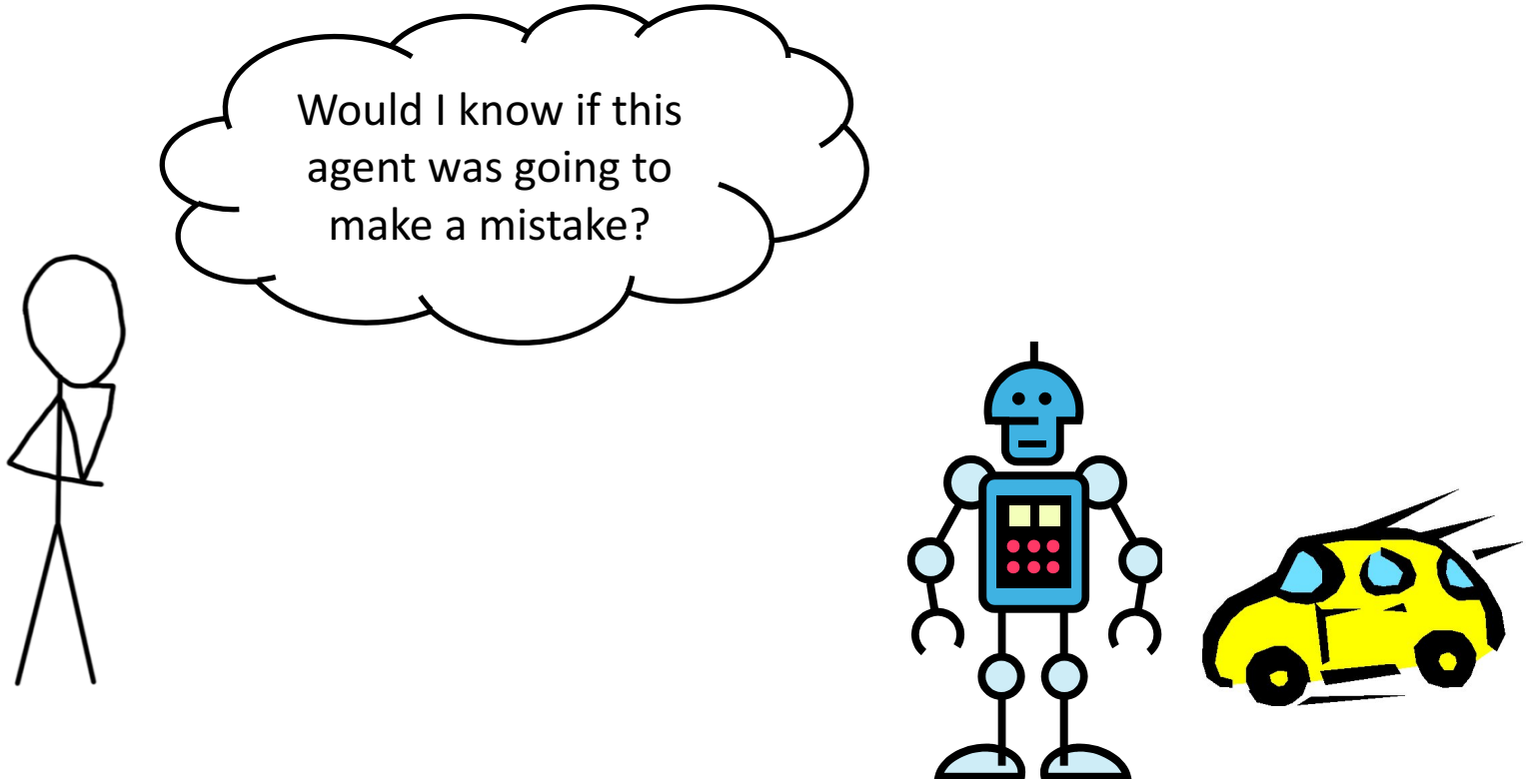
Autonomous Agents



Explaining Agent Behavior

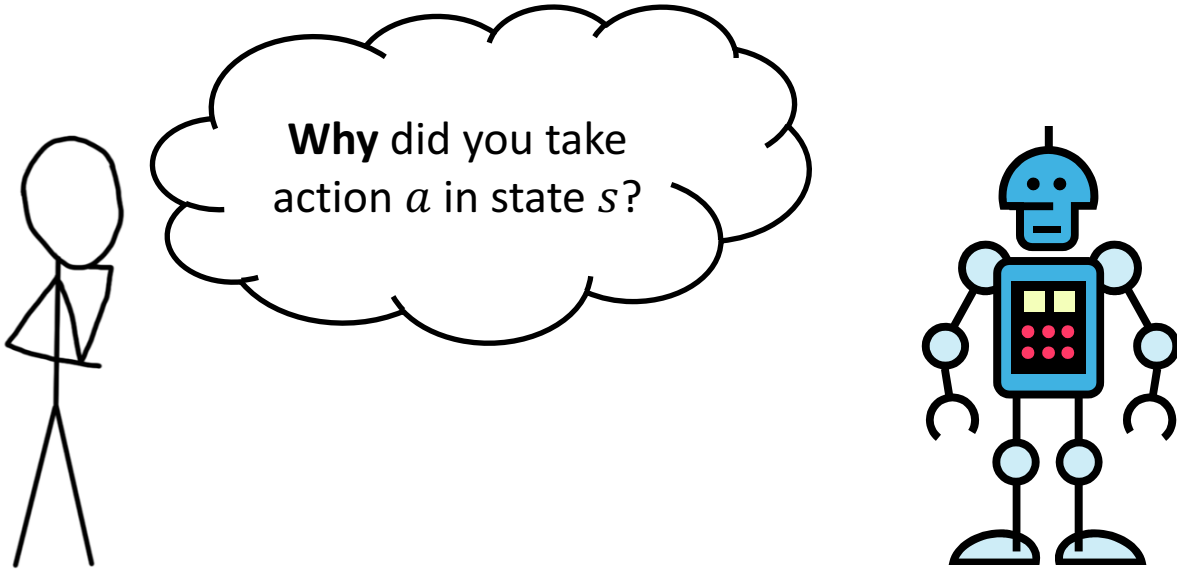


Explaining Agent Behavior



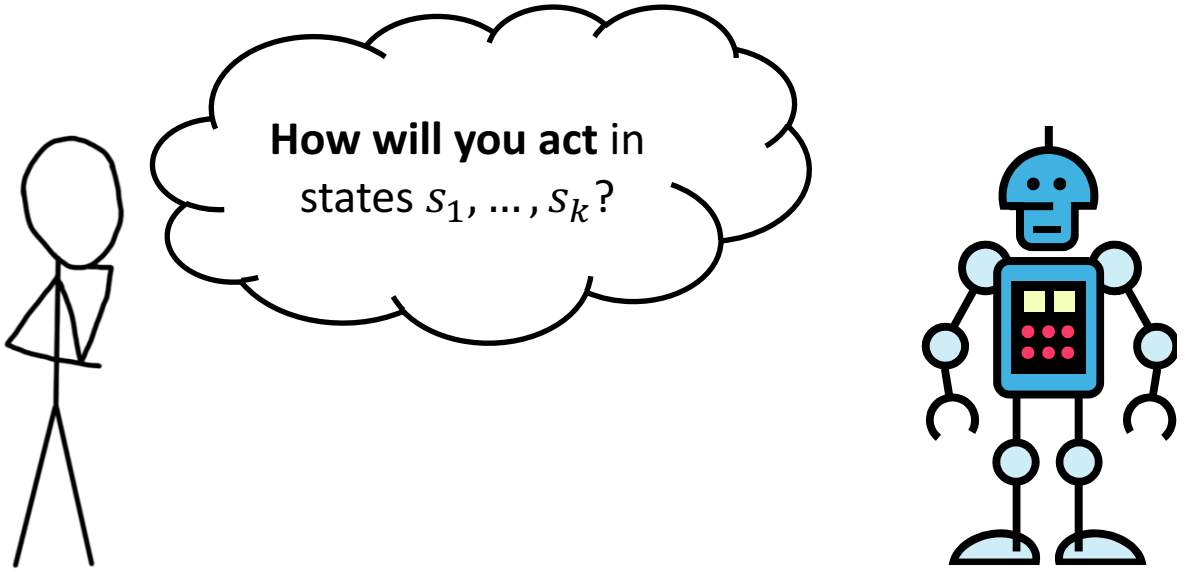
Related Work:
Explain Specific Agent Decisions

- "Explaining robot actions." Lomas et al., 2012.
- "A natural language argumentation interface for explanation generation in Markov decision processes." Dodson et al., 2011
- "Do you get it? user-evaluated explainable bdi agents." Broekens et al., 2010.
- …



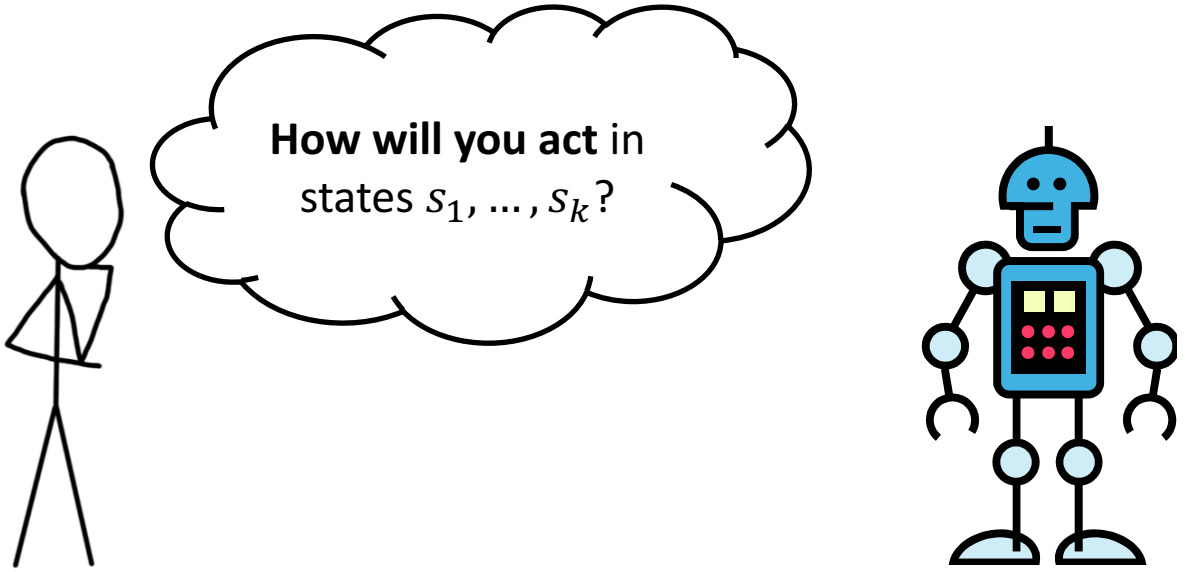
Related Work:
Describe Global Agent Behavior

- Heuristics:
  - "Highlights: Summarizing agent behavior to people." Amir and Amir, 2018
- Models of Humans:
  - "Enabling robots to communicate their objectives." Huang et al. 2017.



Related Work:
Describe Global Agent Behavior

- Heuristics:
  - "Highlights: Summarizing agent behavior to people." Amir and Amir, 2018
- **Models of Humans:**
  - **"Enabling robots to communicate their objectives." Huang et al. 2017.**

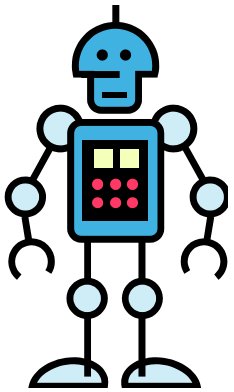


Outline

- Policy Summarization
- Models of Humans
- Contributions
- Simulation Experiment
- Human Subject Experiment
- Open Challenges

Policy Summary

- A set of state-action pairs demonstrating the agent's behavior

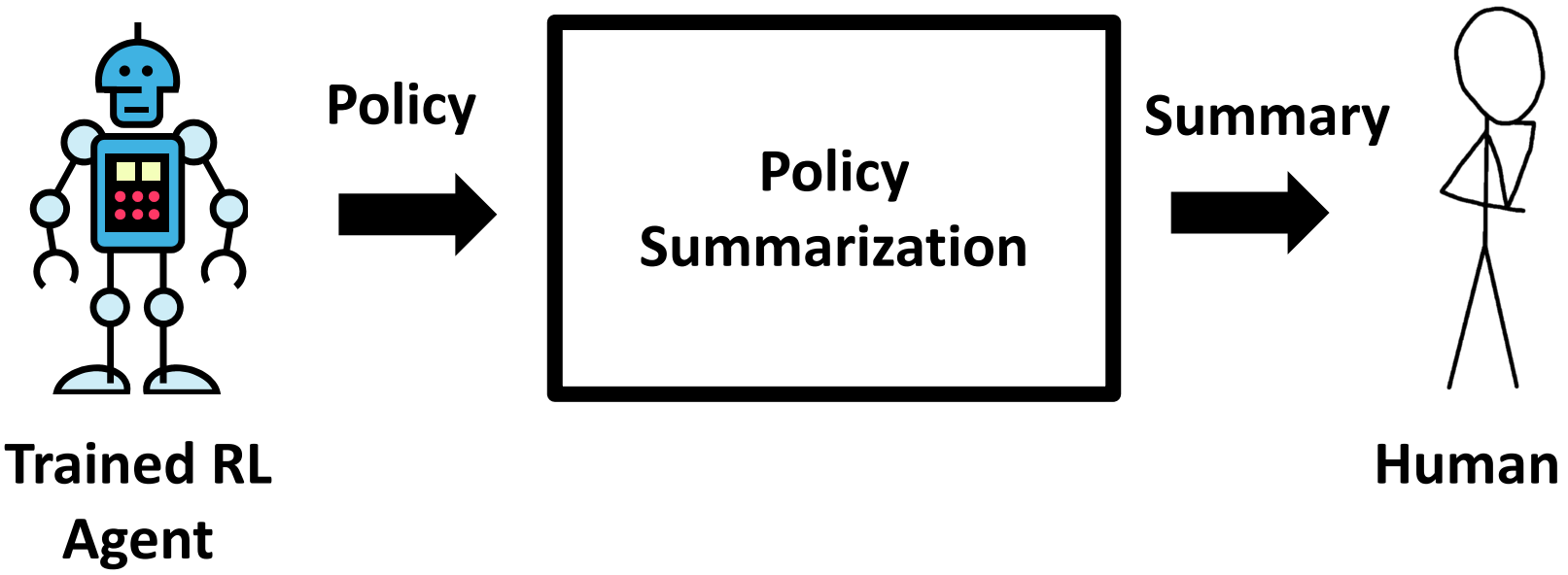


In state s_1, I take action a_1
In state s_2, I take action a_2
…
In state s_k, I take action a_k

k is small to limit human effort

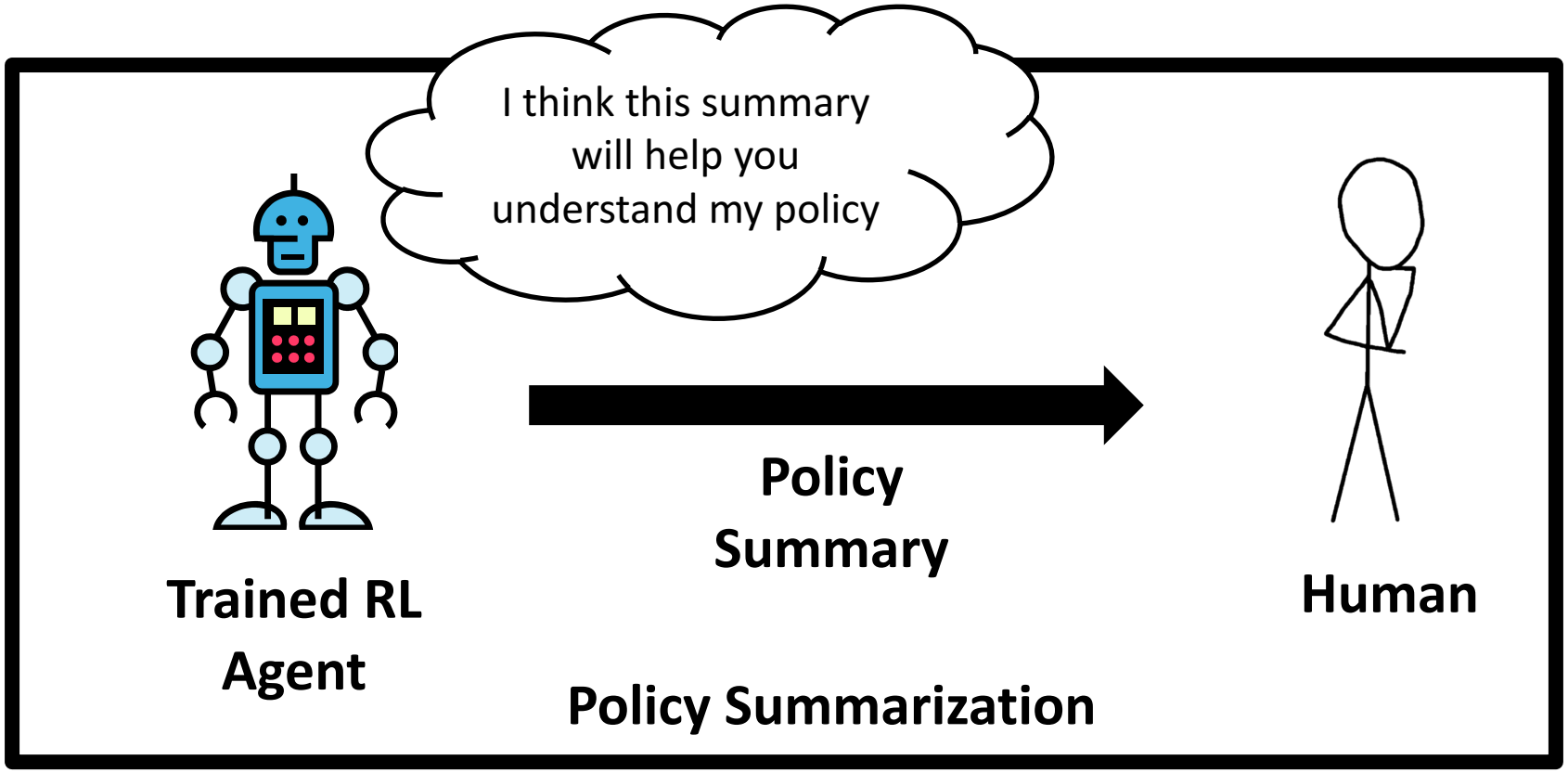
Policy Summarization

- Goal: Describe the global behavior of an agent to a human



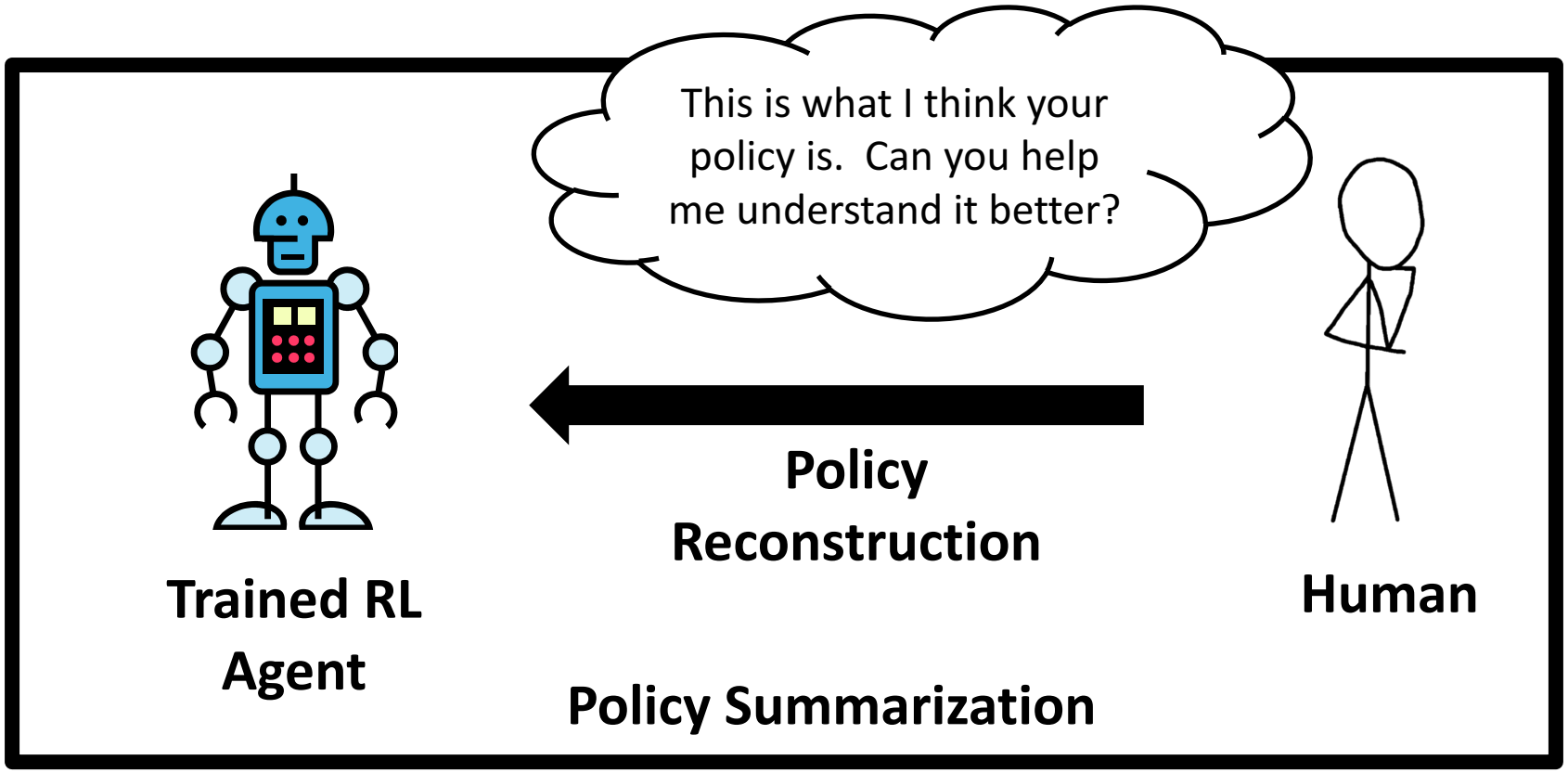
Policy Summarization

- Goal: *Optimize* summary to describe agent behavior to *human*



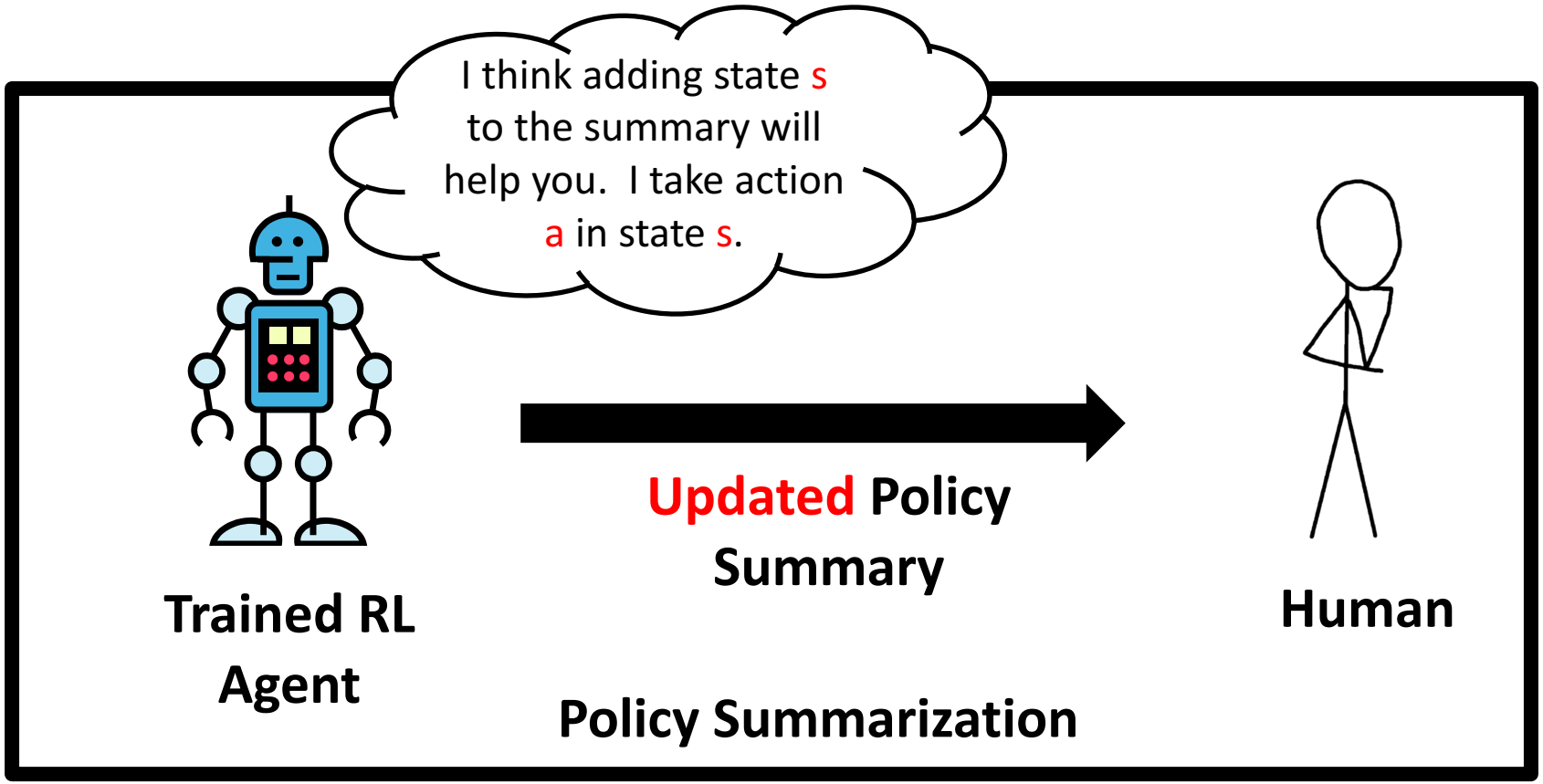
Policy Summarization

- Goal: *Optimize* summary to describe agent behavior to *human*



Policy Summarization

- Goal: *Optimize* summary to describe agent behavior to *human*

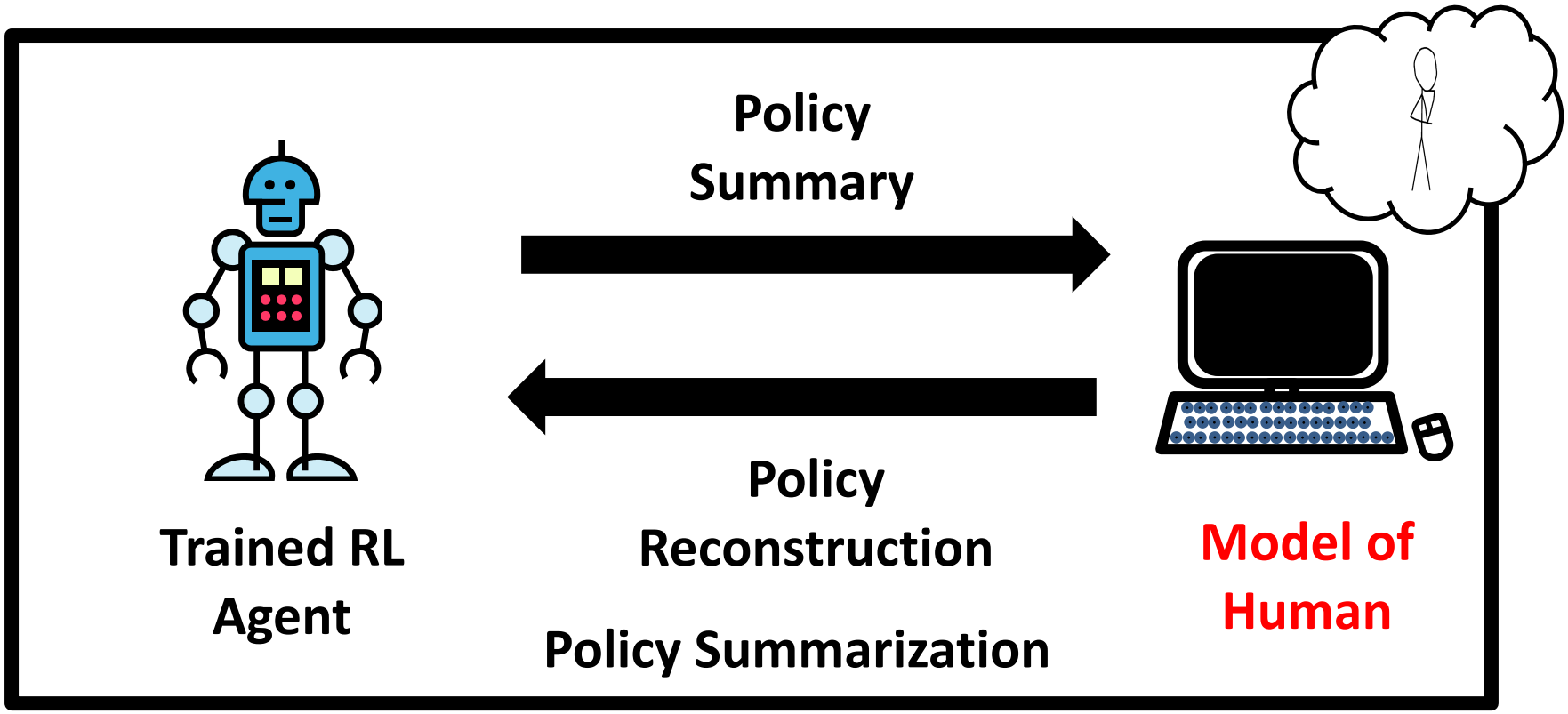


Policy Summarization

- Goal: *Optimize* summary to describe agent behavior to *human*

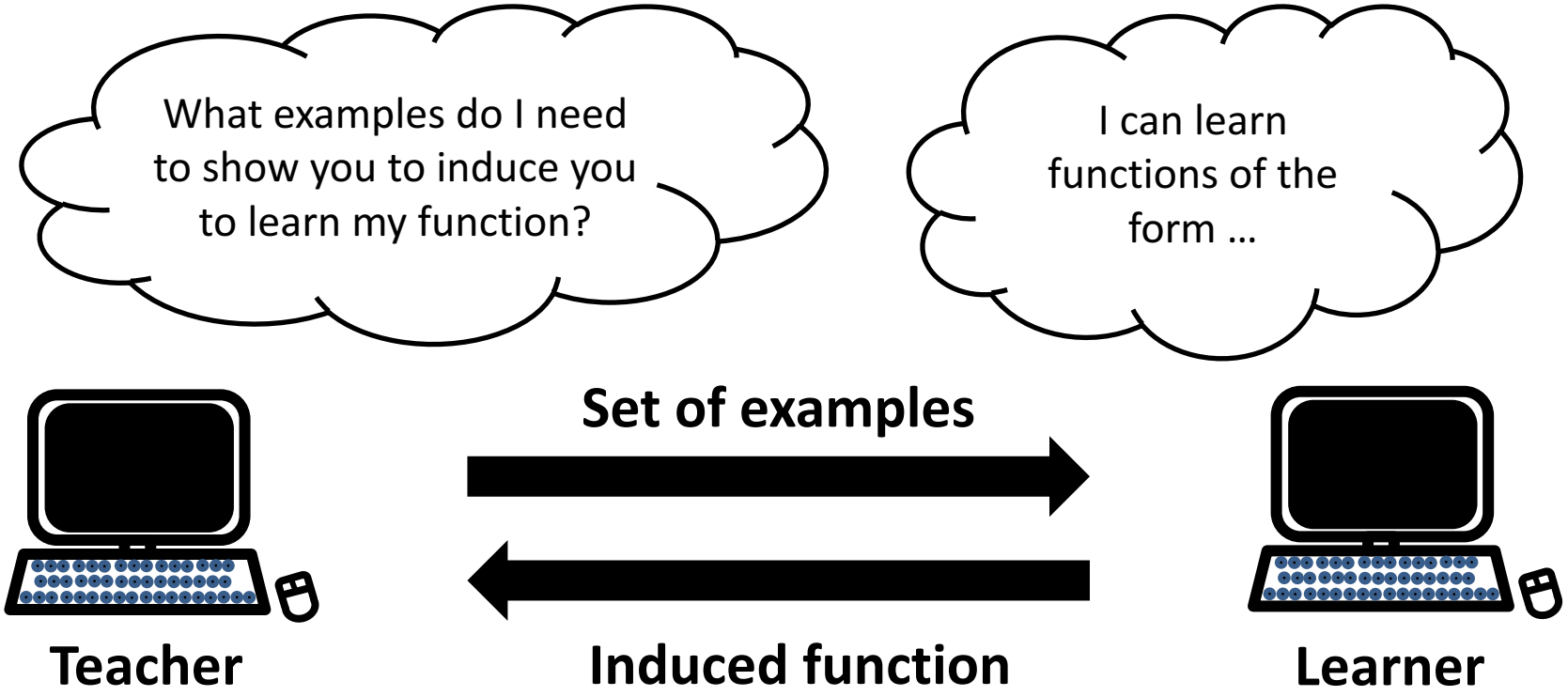
Policy Summarization

- Goal: *Optimize* summary to describe agent behavior to *model of human*



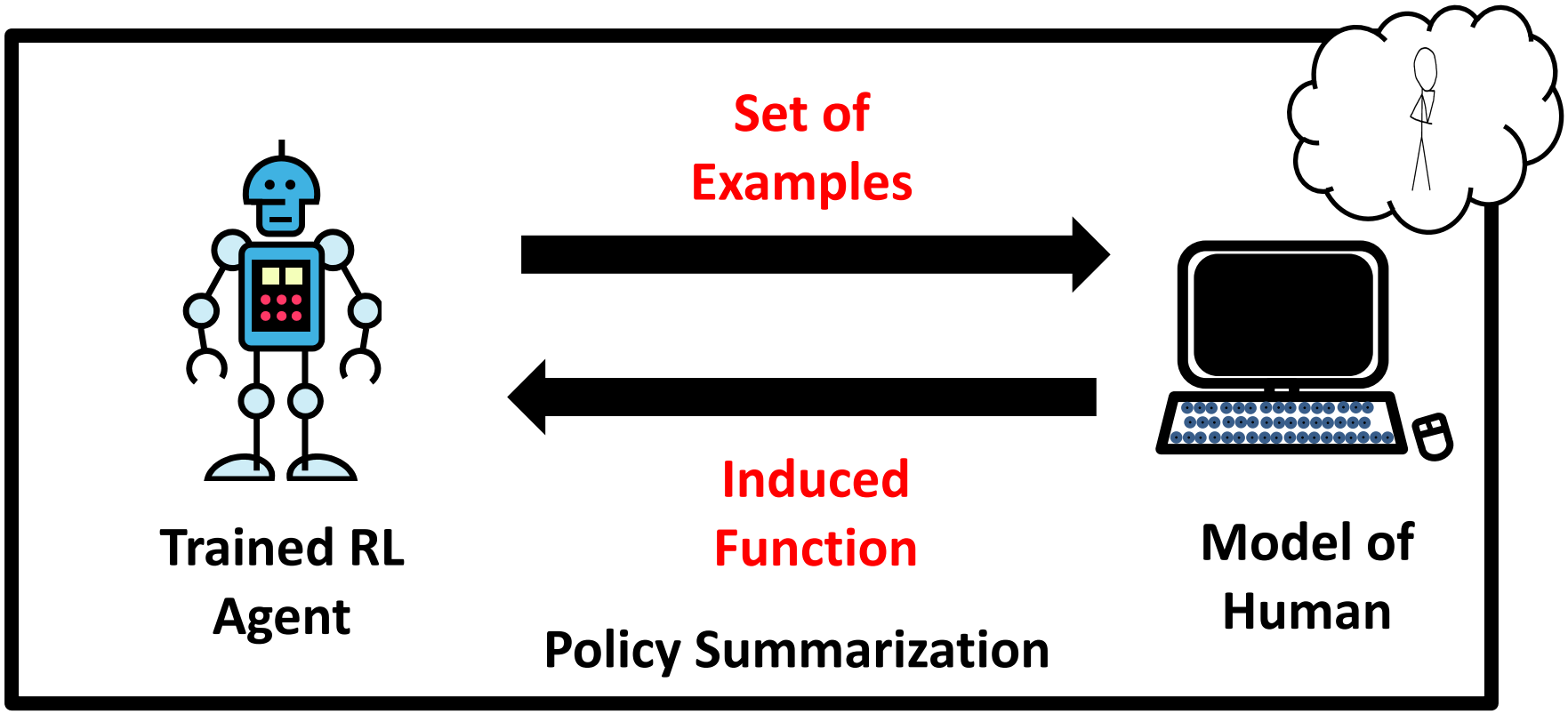
Mechanism: Machine Teaching

- Goal: Optimize a set of examples to induce a known model in a known class of learners.



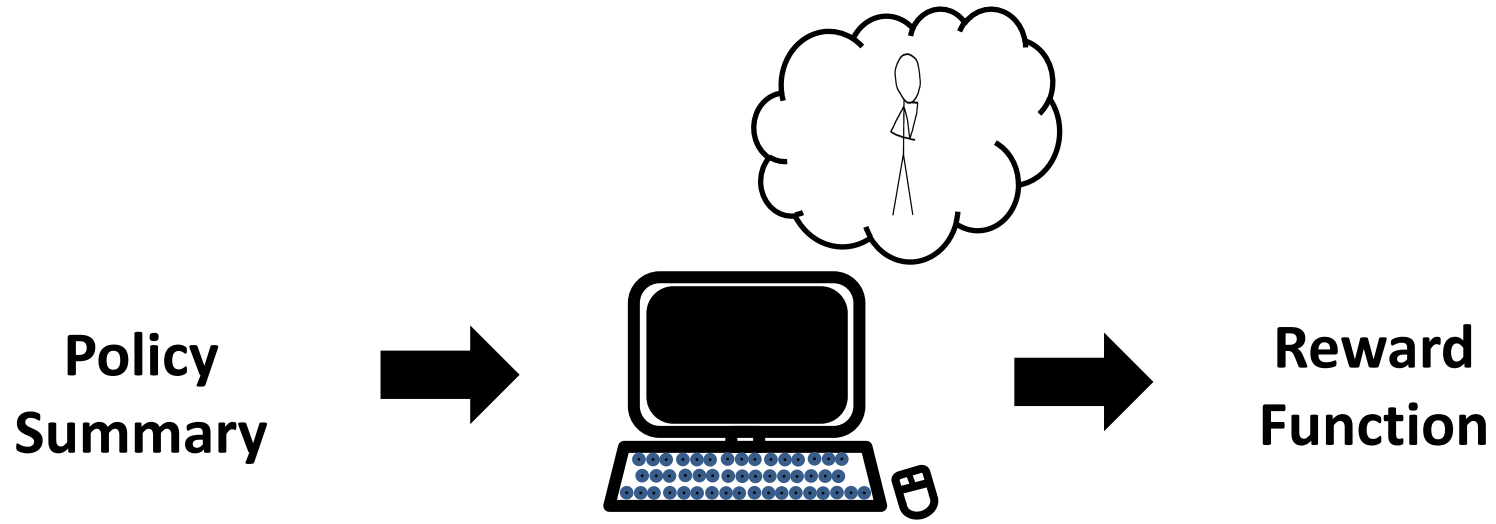
Machine Teaching-Based Policy Summarization

- Goal: *Optimize* summary to describe agent behavior to *model of human*



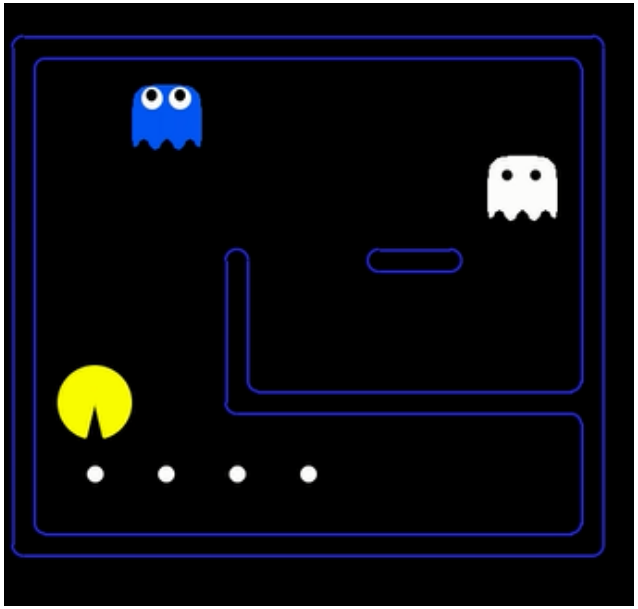
Models of Humans – Inverse RL

- Goal: Extrapolate policy by first learning reward function



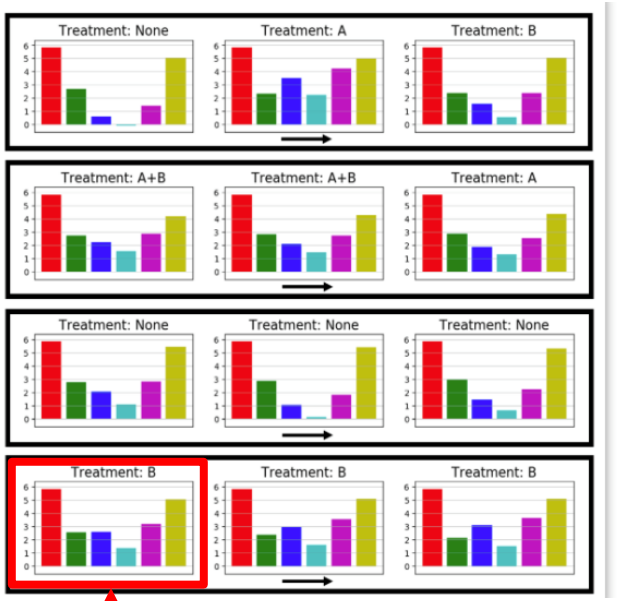
"Enabling robots to communicate their objectives." Huang et al. 2017.

IRL for PAC-MAN



Reward = ↑ pellets + ↓ ghosts

IRL for HIV Simulator

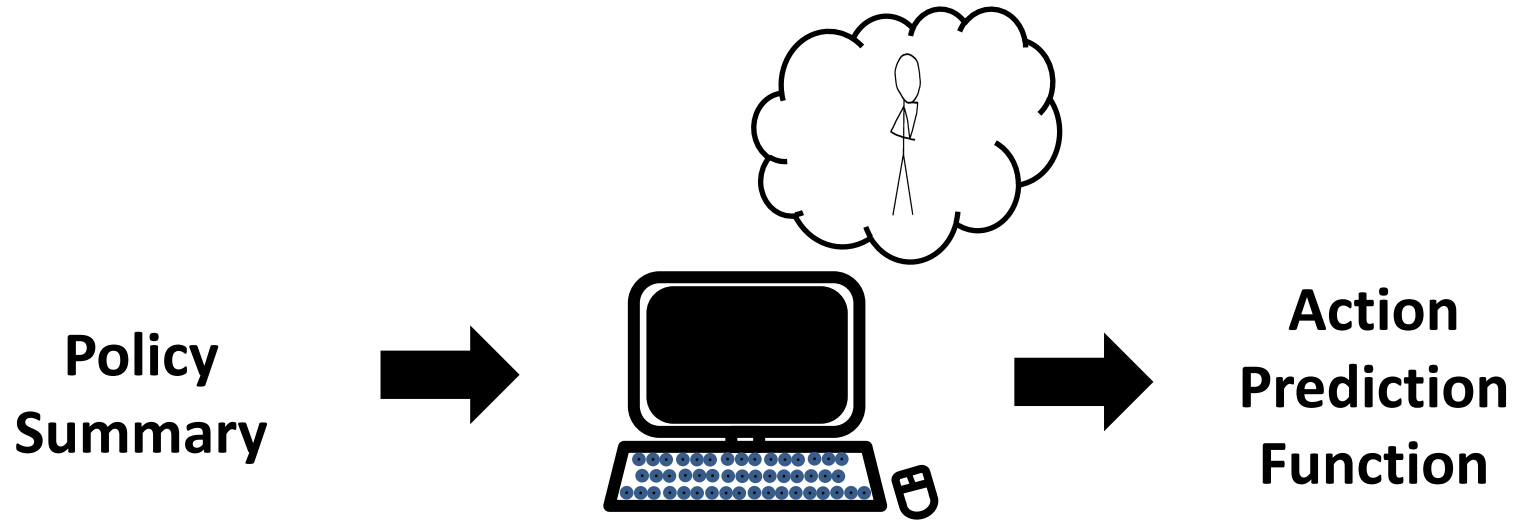


State space

Reward = ???

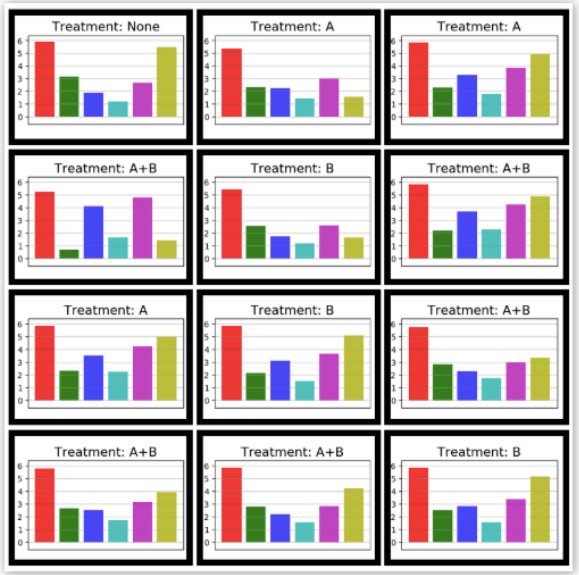
Models of Humans – Imitation Learning

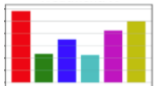
- Goal: Extrapolate policy by predicting the action taken in similar states



Contribution

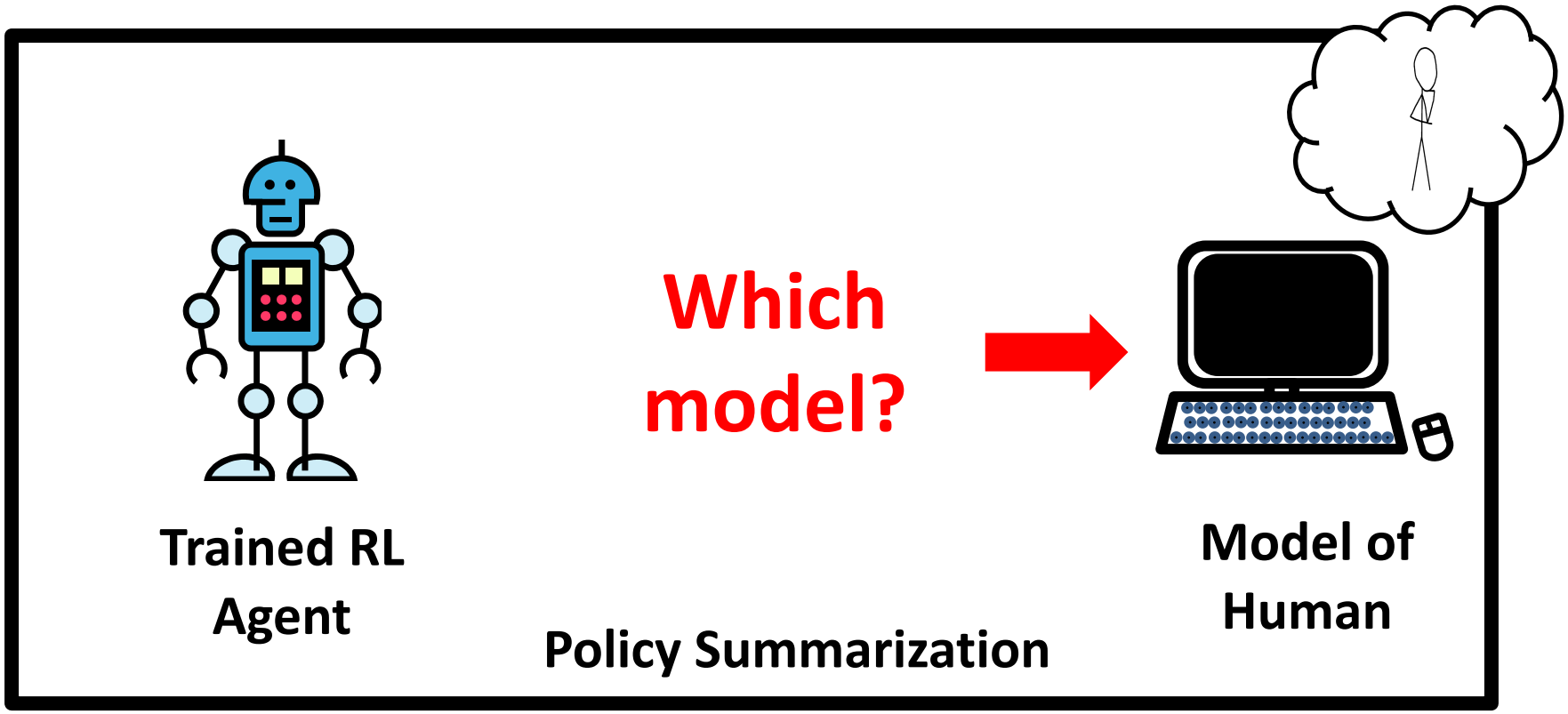
IL for HIV Simulator



In states like  **, it gives Treatment A**

Contributions

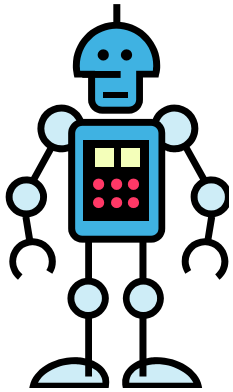
- How should we model humans during policy summarization?



Contributions

- Which reconstruction models do people use in different domains?
- Is it necessary to match the summarization model to peoples' reconstruction model to produce effective summaries?

Evaluation



Trained RL Agent

π

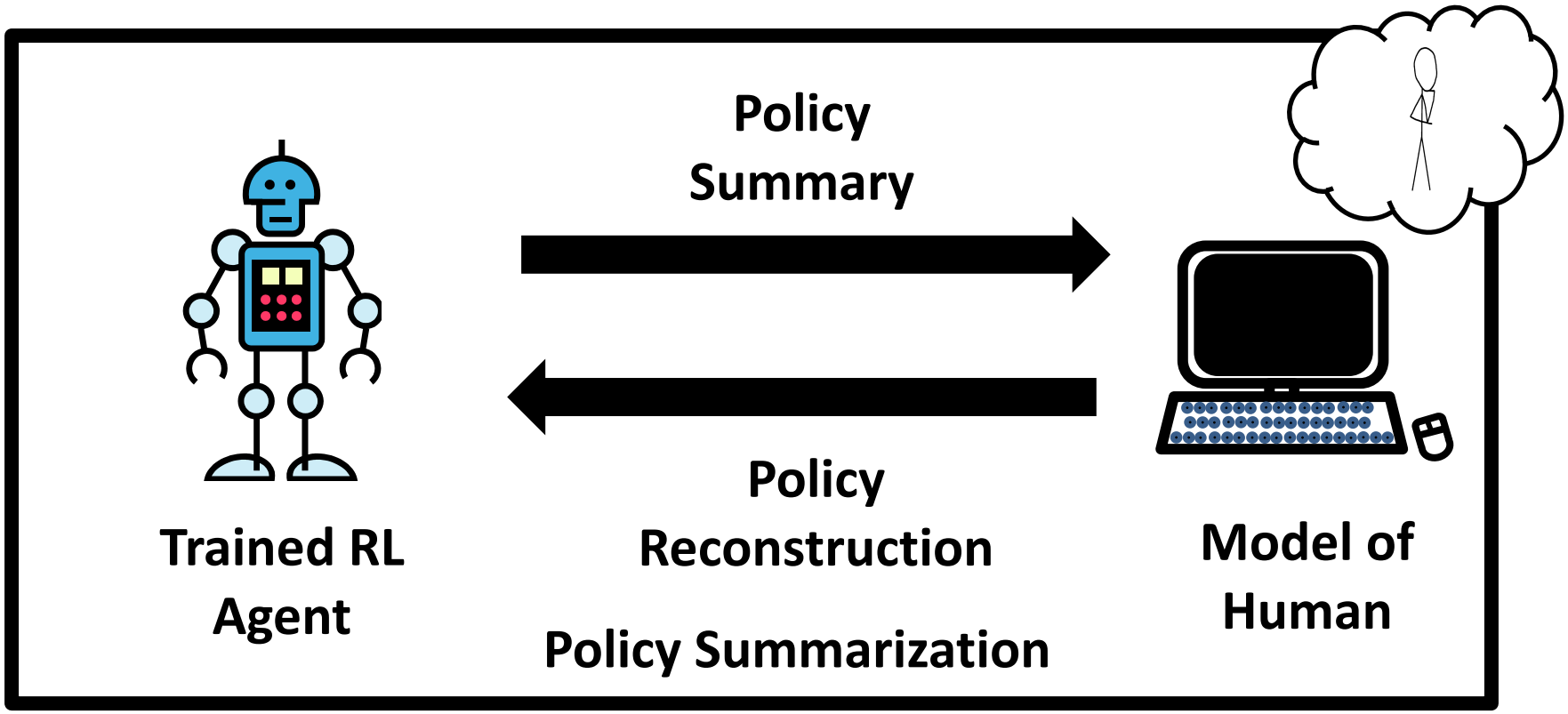
Policy Summarization

Summary

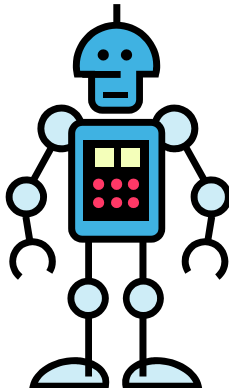
Summarization

Recap: Policy Summarization

- Goal: *Optimize* summary to describe agent behavior to *model of human*



Evaluation



Trained RL Agent

π

Policy Summarization

Summary

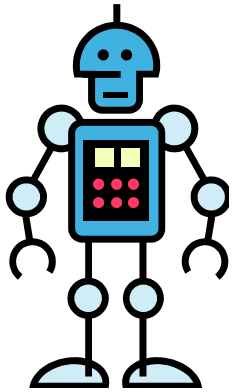
Human

$\hat{\pi}$

Summarization

Reconstruction

Evaluation



Trained RL Agent

π

Policy Summarization

Summary

Human

$\hat{\pi}$

Quality of Summary:

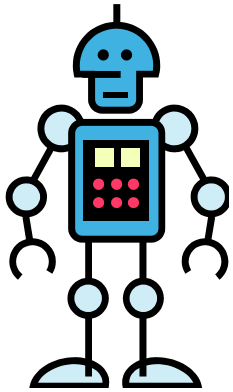
- Accuracy($\hat{\pi}$, π)
- Value(π) – Value($\hat{\pi}$)

Summarization **Reconstruction** **Evaluation**

Simulation Experiment

- Is it necessary to match the summarization model to the reconstruction model to produce high-quality reconstructions?

Simulation Experiment



Trained RL Agent

π

Imitation Learning

Summary

Inverse RL

Model

$\hat{\pi}$

For example:

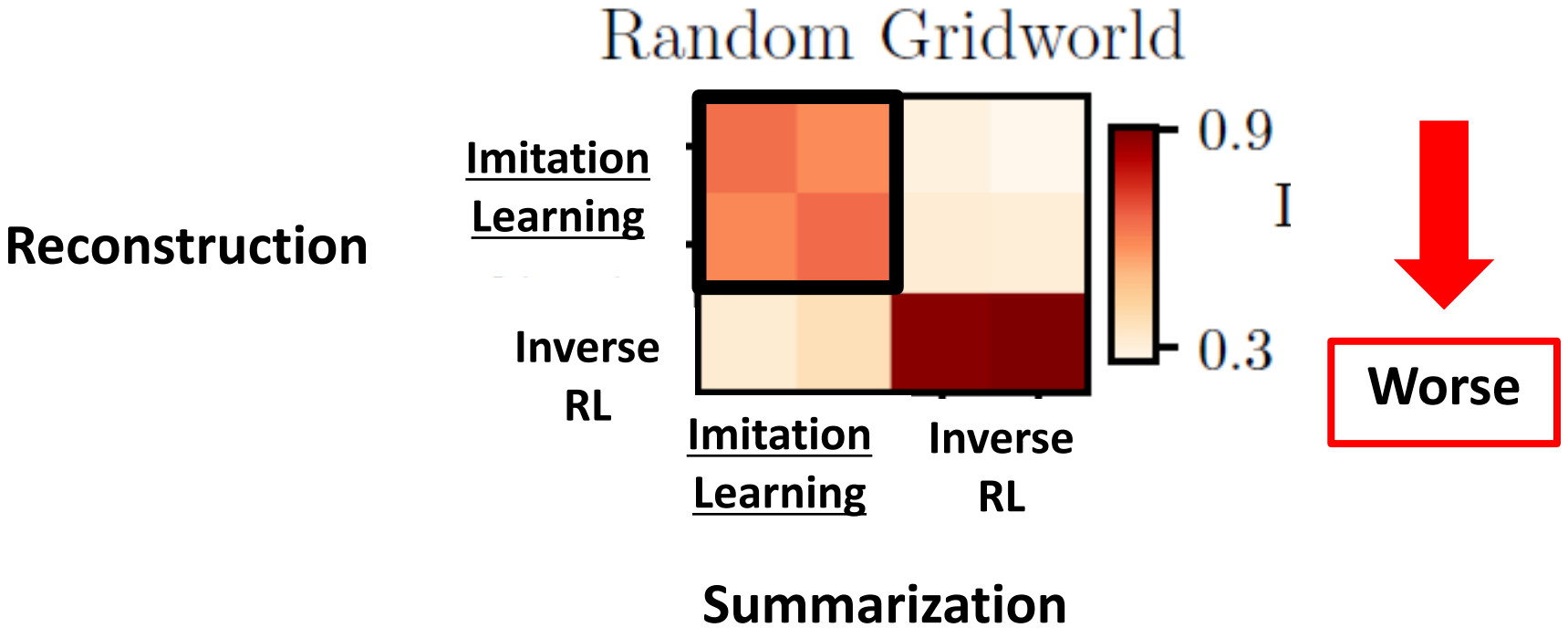
Quality of Summary:

- Accuracy($\hat{\pi}$, π)
- Value(π) – Value($\hat{\pi}$)

Summarization **Reconstruction** **Evaluation**

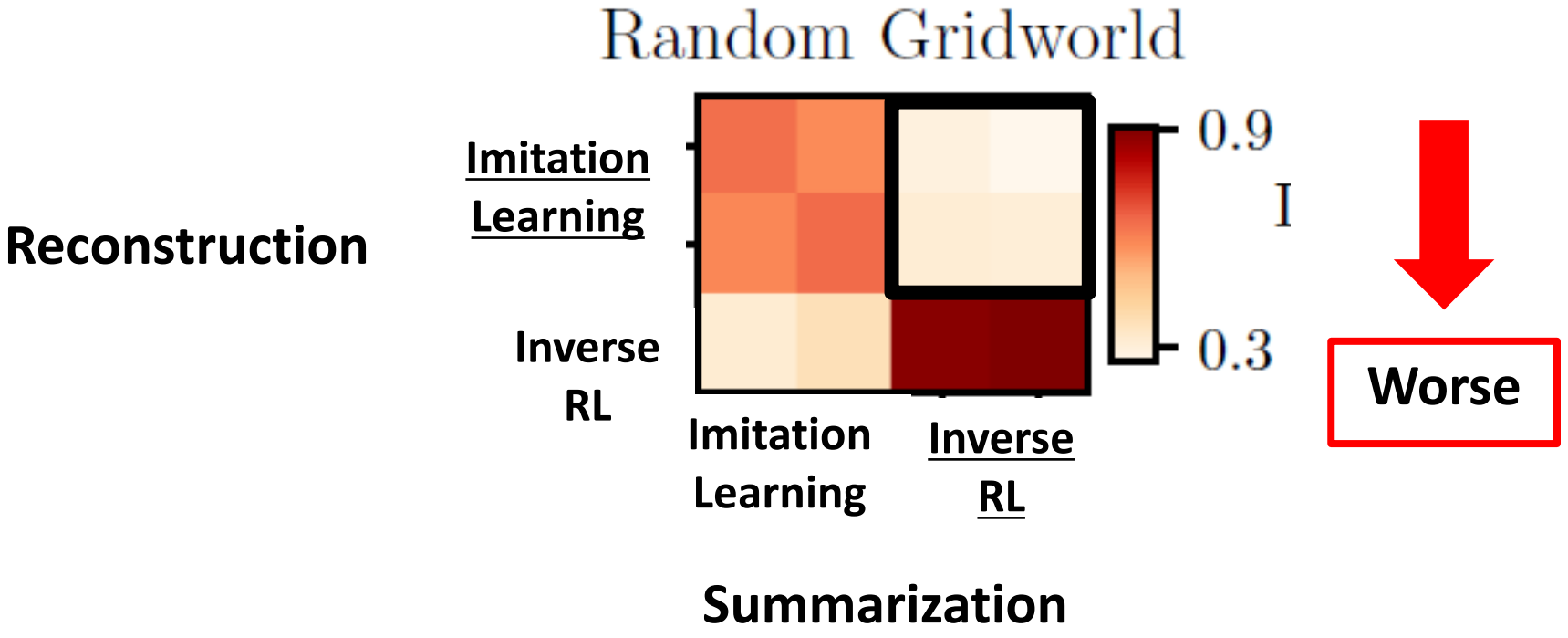
Simulation Experiment

- Matched models: high policy-reconstruction quality



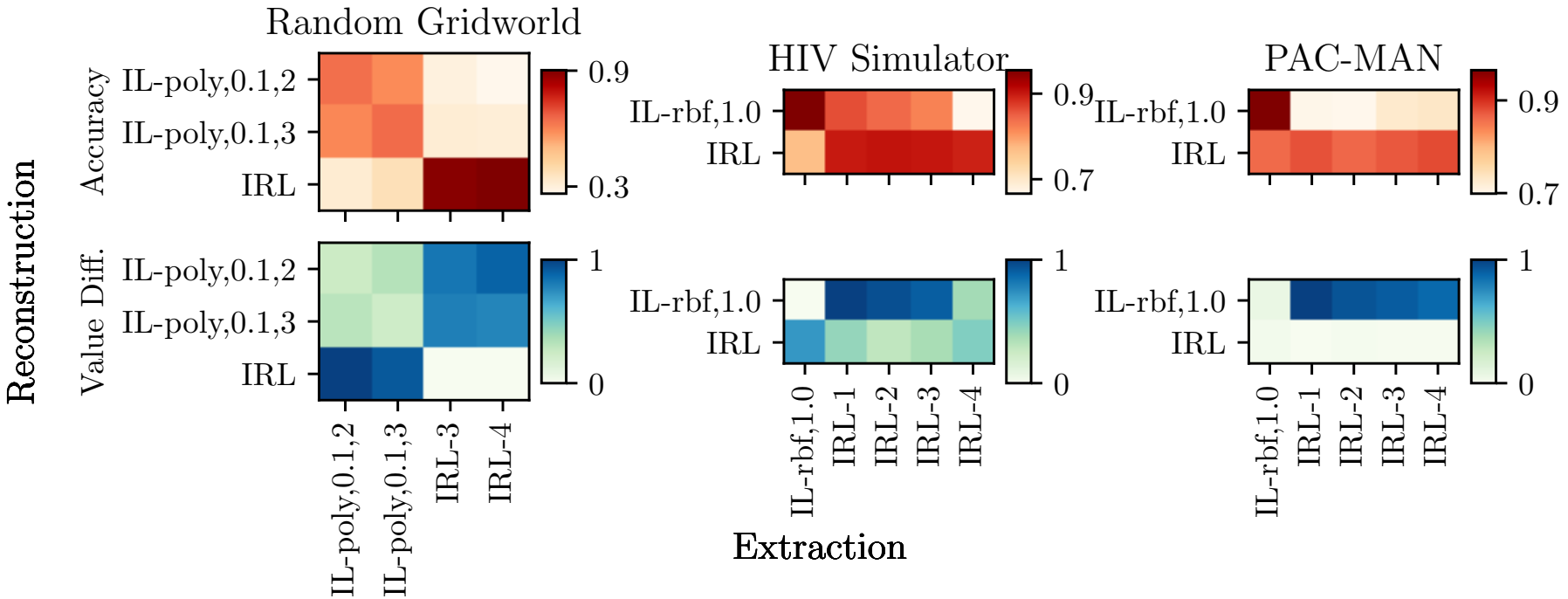
Simulation Experiment

- Mismatched models: low policy-reconstruction quality



Simulation Experiment

- Policy-reconstruction quality is worse when models don't match



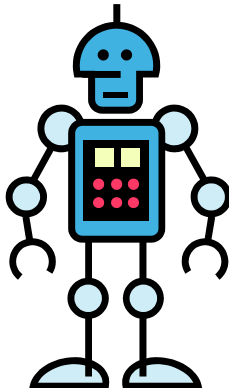
Simulation Experiment

- Assuming a different model than the reconstruction model during summarization results in worse policy reconstructions.
- We shouldn’t always use IRL in summarization if people sometimes do IL.

Human Subject Experiment

- Which reconstruction models do people use?
- Does our finding about the importance of matching summarization models to reconstruction models hold for humans?

Evaluation



Trained RL Agent

π

Imitation Learning

Summary

Human

$\hat{\pi}$

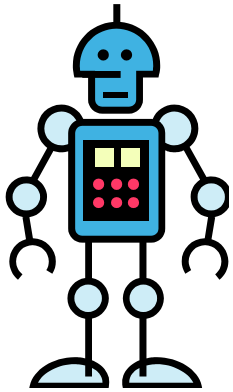
For example:

Quality of Summary:

- Accuracy($\hat{\pi}$, π)

Summarization **Reconstruction** **Evaluation**

Evaluation



Trained RL Agent

π

Imitation Learning

Summary

Reconstruction model = ???

Human

$\hat{\pi}$

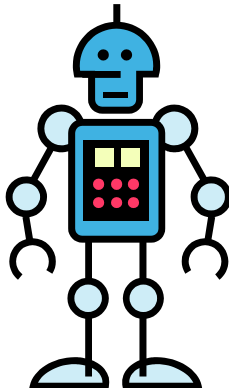
For example:

Quality of Summary:

- Accuracy($\hat{\pi}$, π)

Summarization **Reconstruction** **Evaluation**

Evaluation



Trained RL Agent

π

Imitation Learning

Summary

Human

$\hat{\pi}$

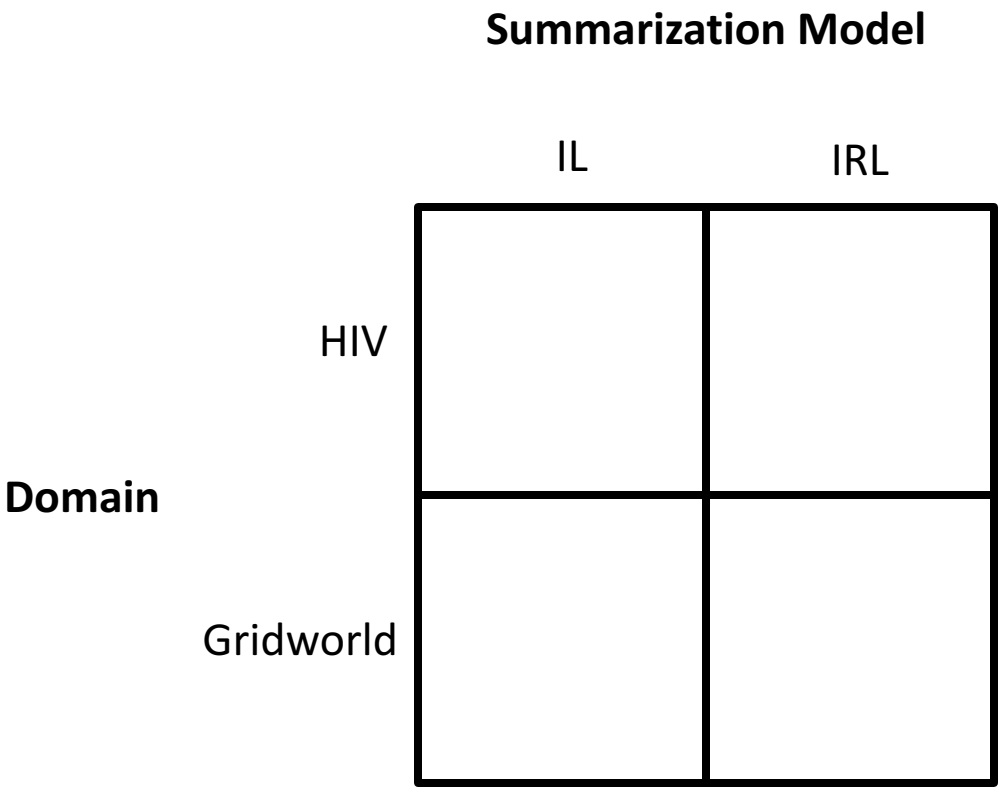
For example:

Quality of Summary:

- Accuracy($\hat{\pi}$, π) **for a sample of s**

Summarization **Reconstruction** **Evaluation**

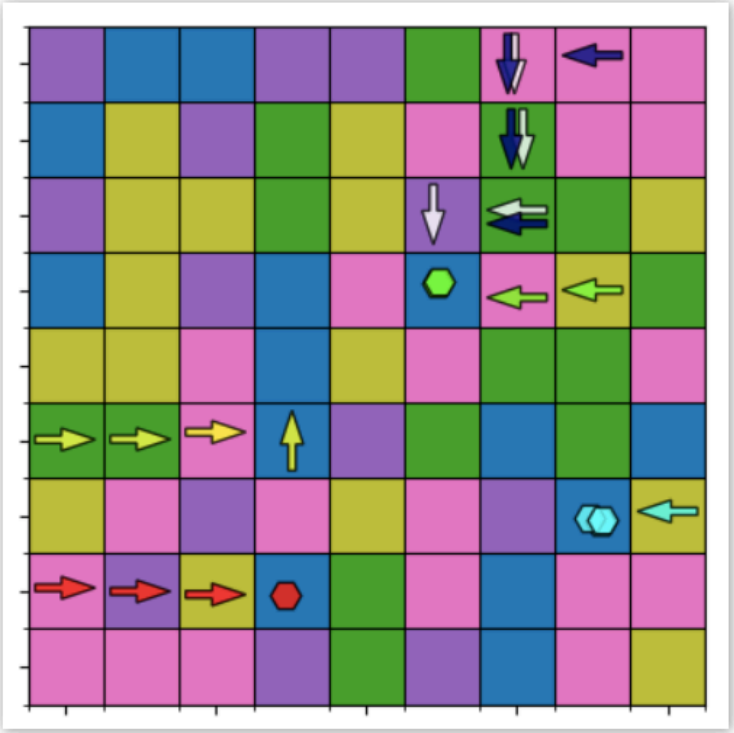
Human Subject Experiment



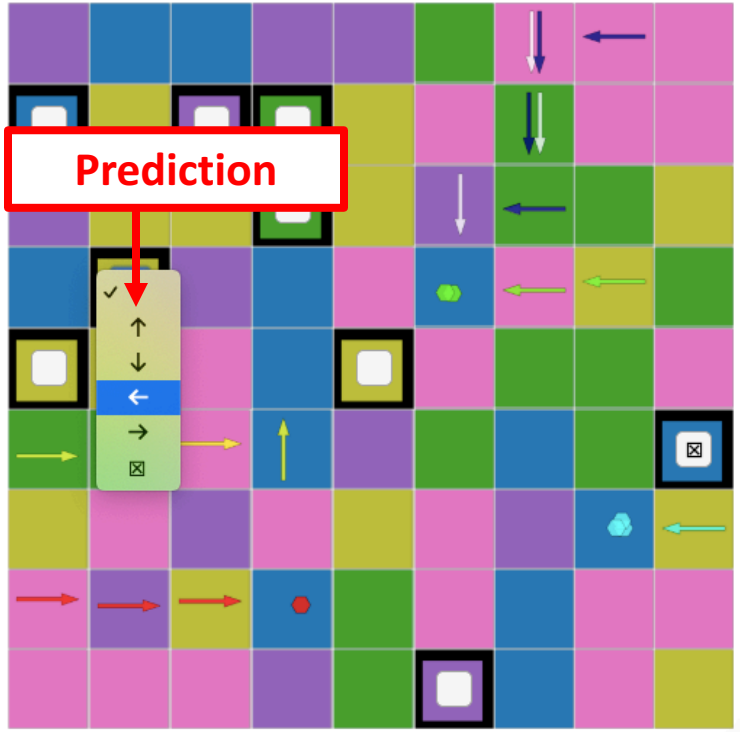
Human Subject Experiment

- Task: predict the agent's behavior in a selection of states

Summary

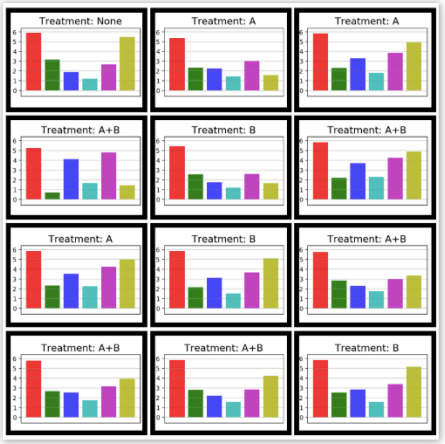


Task



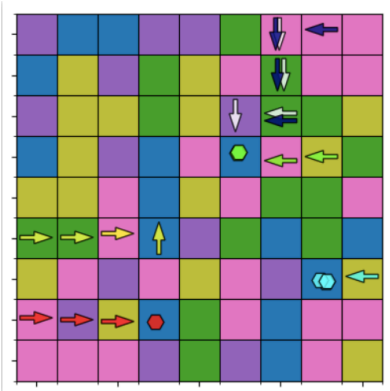
Qualitative Reconstruction – HIV

- People reconstruct with IL in the HIV domain
  - **IL-based reconstruction: 78%**
    - E.g. "*I chose based on the similarity of the blood tests levels from the scenarios on the left*."
  - **IRL-based reconstruction: 1 person**
    - E.g. "*Treatment A is used to decrease middle ones(blue, light blue and purple)…*"



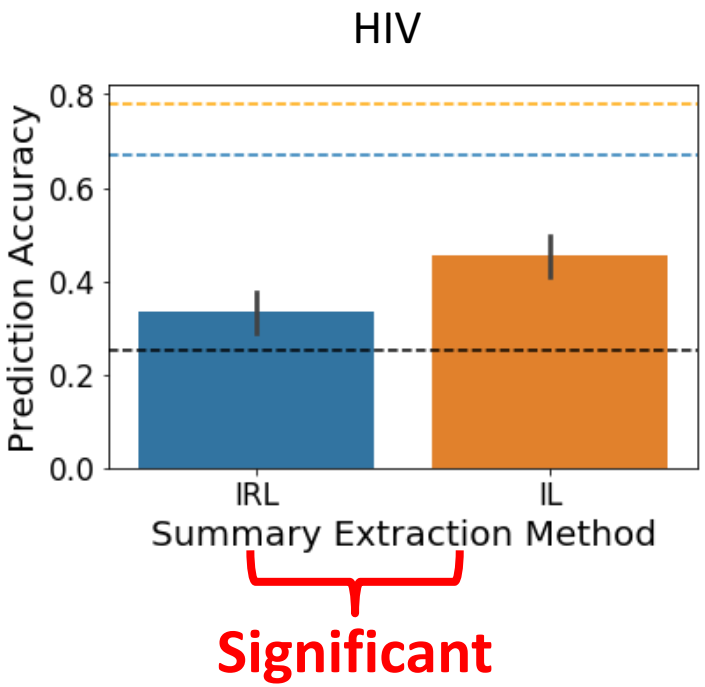
Qualitative Reconstruction – Gridworld

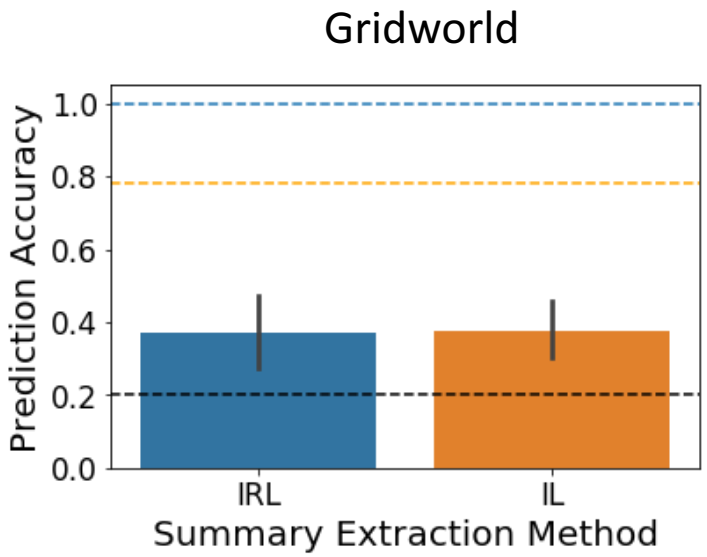
- People reconstruct with IRL in the Gridworld
  - **IL-based reconstruction: 15%**
    - E.g. "*I tried comparing the colors and deciding which action was more frequent for a color*."
  - **IRL-based reconstruction: 27%**
    - E.g. "*I decided that the computer seems to be always working towards a blue square. I chose the simplest path to get to a blue square*"



Quantitative Reconstruction

- In the HIV domain, people reconstructed the policy more accurately with the matched IL summary
- In the gridworld domain, it made no difference



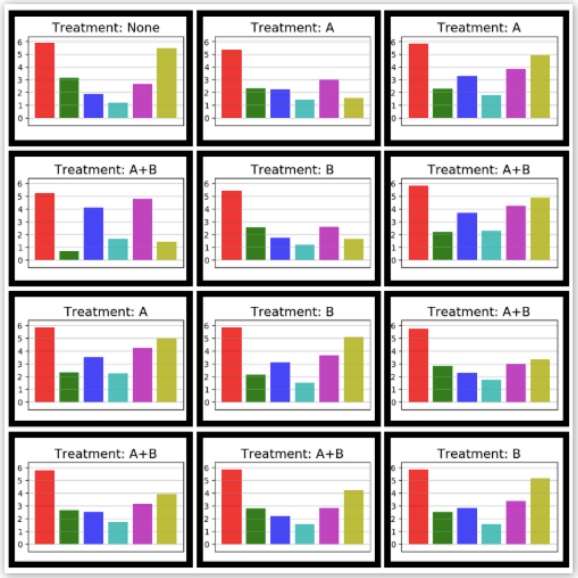


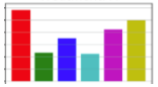
Human Subject Experiment

- People use different models in different domains; it is important to consider IL as a possible model.
- Mismatch between the summarization model and peoples' reconstruction model can result in worse policy reconstructions.

Our Contributions

- Introduced an imitation learning model for policy summarization



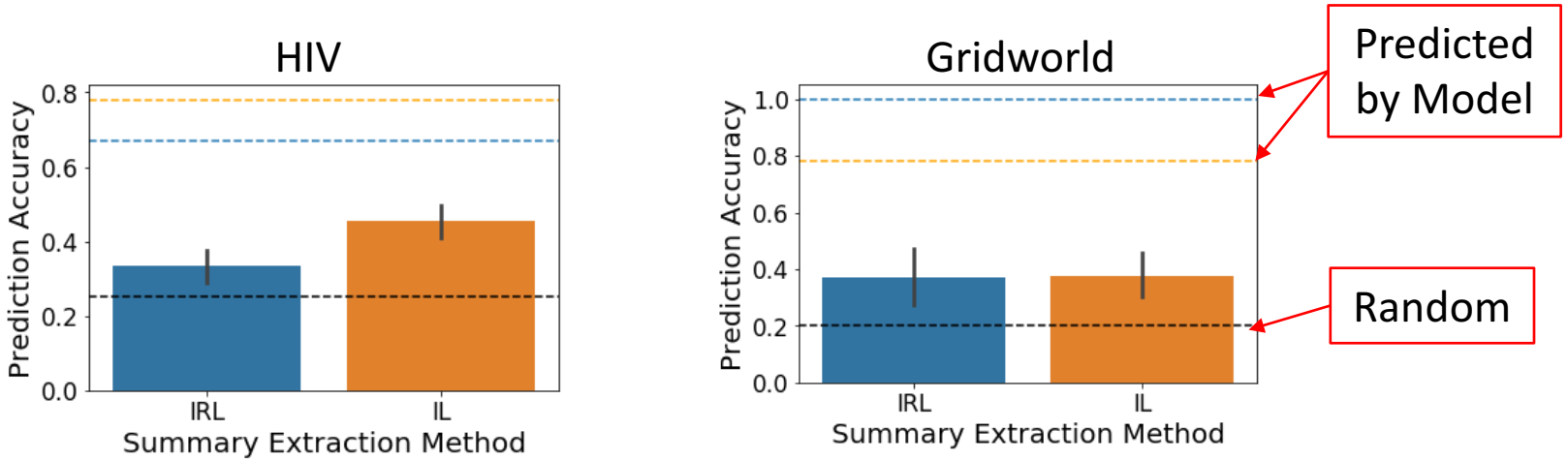
In states like  **, it gives Treatment A**

Our Contributions

- Results:
  - Assuming a different model than the reconstruction model during summarization results in worse quality policy-reconstructions
  - People use different models in different domains; it is important to consider IL as a possible model

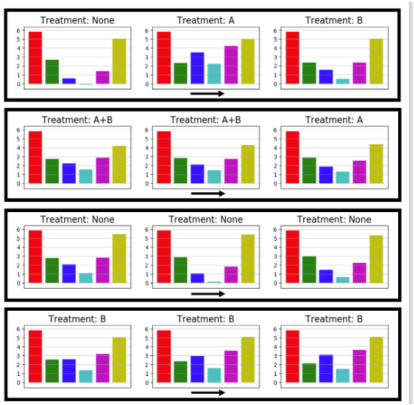
Open Challenges

- Can we develop better models of how humans reconstruct policies?
  - Reconstruction was better than random but worse than predicted
    - Better similarity metrics?
    - Sparsity assumptions in IRL?



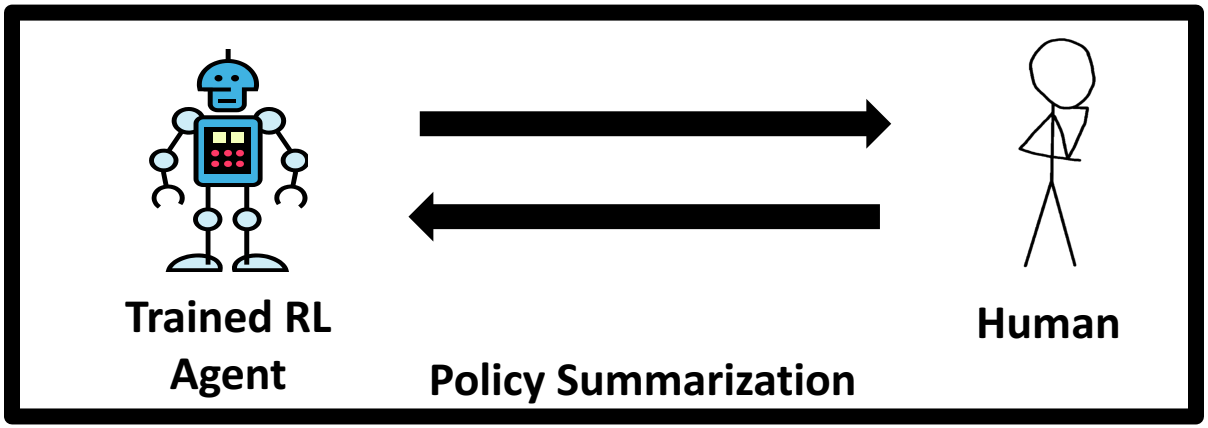
Open Challenges

- How to scale these approaches to high dimensions?
  - People can only view a few dimensions
    - Can we develop better visualizations for high-dimensional data?
    - Can we build which dimensions to show into the models?



Open Challenges

- Can we elicit peoples' reconstruction model?
  - Knowing reconstruction models is important
    - Can we discover properties of which domains correspond to which models?
    - Can we use human-in-the-loop approaches to learn peoples' reconstruction model during summarization?



Open Challenges – Recap

- More nuanced models of humans
- Visualizing high-dimensional states
- Eliciting domain-specific reconstruction models through human-in-the-loop approaches