IMPERIAL

Insights from the Inverse: Leveraging Inverse Reinforcement Learning for Actionable Decision-Making in Healthcare

Sonali Parbhoo, PhD.

Imperial College London

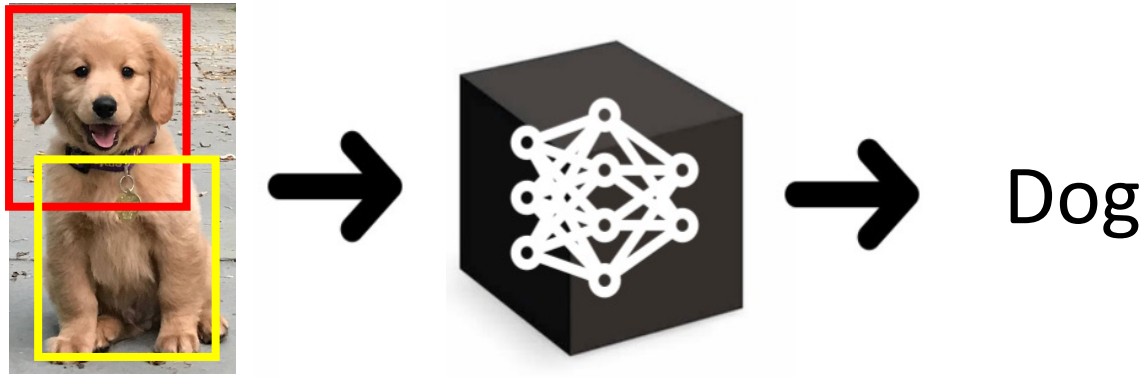
Apple Healthcare Workshop

March 29, 2024

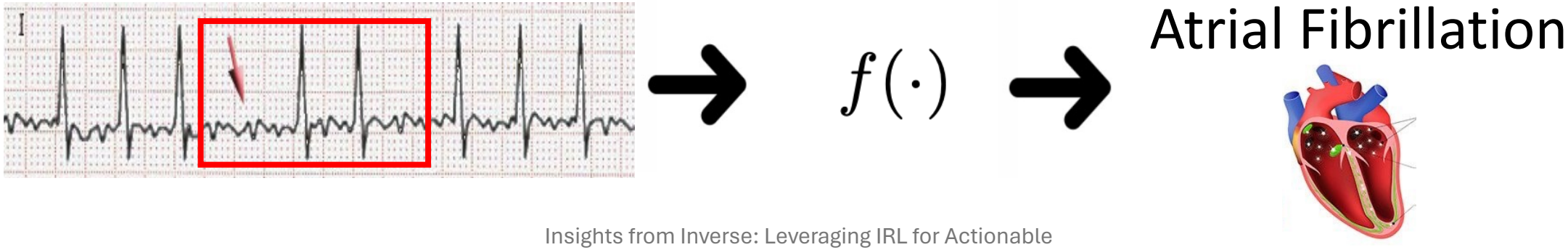
Collaborators: Leo Benac, Abhishek Sharma, Finale Doshi-Velez

Forward problems are relatively straightforward to solve in practice

E.g. Given an image, it is easy to identify elements of the image for classification.

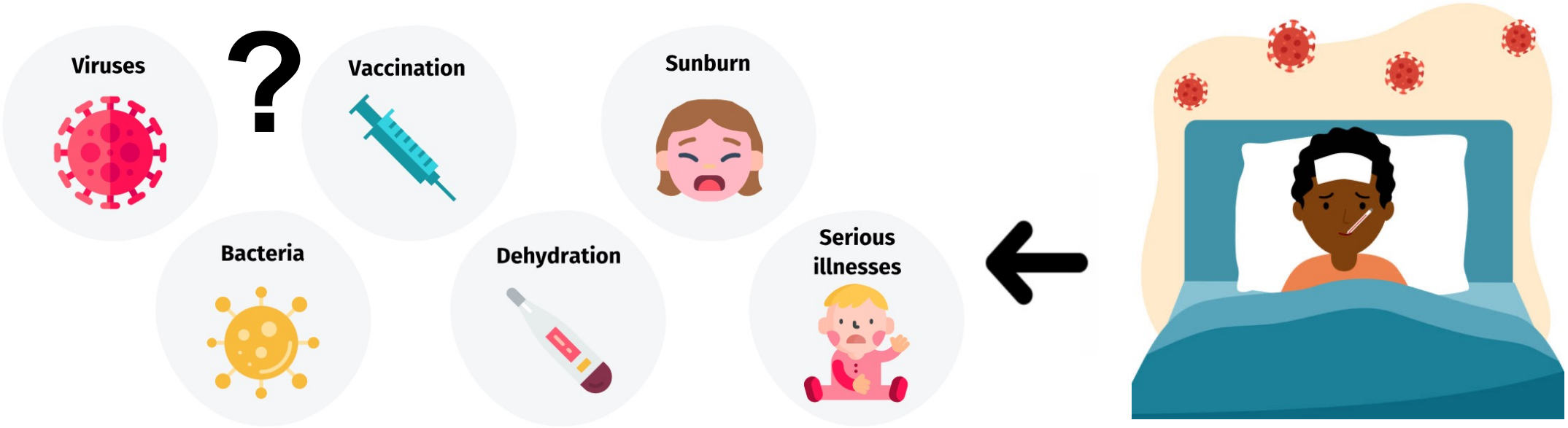


Given an ECG, it is relatively straightforward to characterise some diseases



Inverse problems are very common, but much harder to solve

Many problems in healthcare can be thought of as inverse problems.



How did we get there?

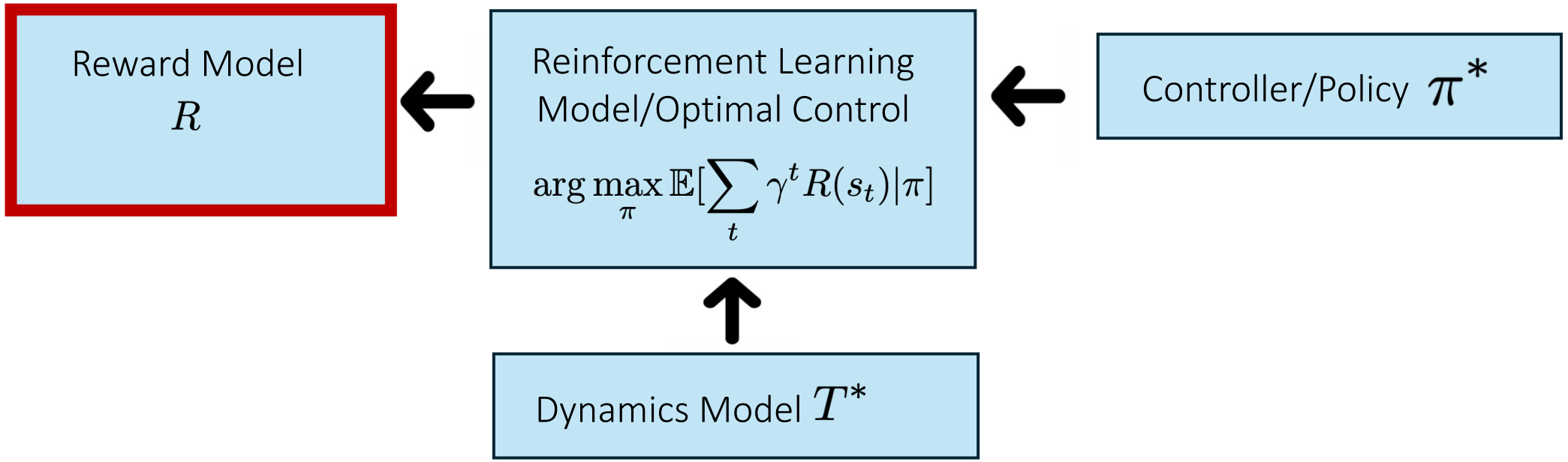
Shivam has a fever

These are much harder to solve because many possible solutions may exist.

Inverse Reinforcement Learning

Goal: Determine the **intent** of a human expert i.e. find the reward model

$R(s,a) = \mathbb{E}[r|s,a]$ such that $\pi^* = \arg\max_\pi \sum \gamma^t \mathbb{E}_\pi[\mathbb{E}[r_t|s_t,a_t]]$



IRL may end up choosing **very different actions** to an expert to achieve the same result

Inverse Reinforcement Learning does not tell us the full story

Determining the intent of an expert using IRL may not capture the nuances in decision-making.

A



B



Low Blood Pressure

High Heart Rate

Low Blood Oxygen

Fever



Expert Policy

Give vasopressors and antibiotics

→

A

Full recovery after 72 hrs

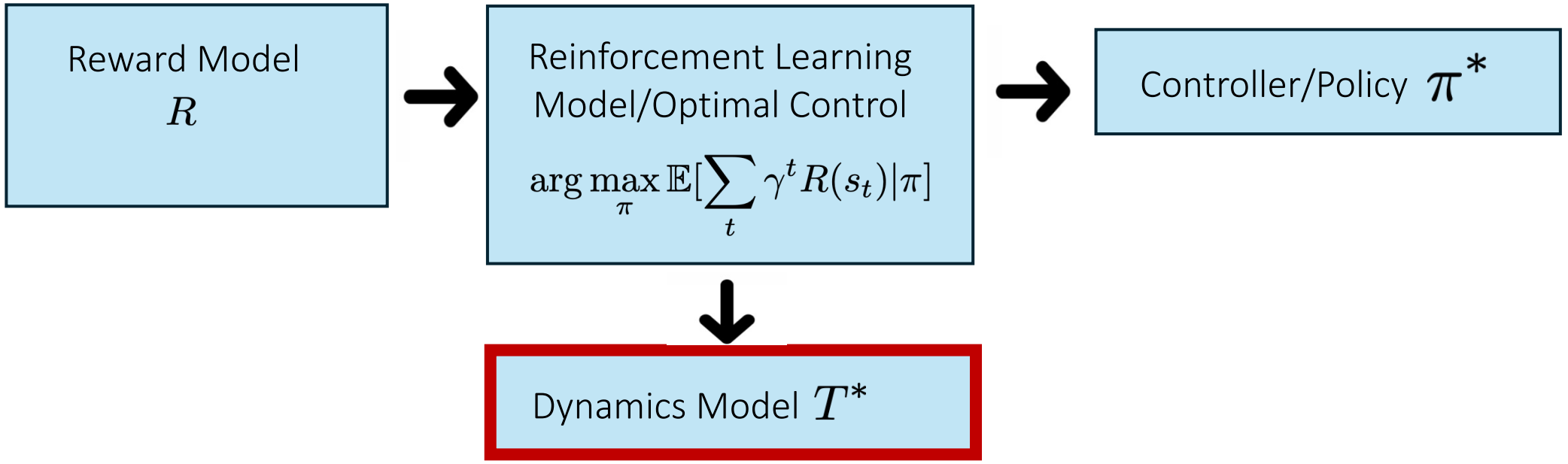
B

Full recovery after 72 hrs but experienced several complications at 48 hrs

Which of these scenarios would you prefer?

Our contribution: Inverse Transition Learning

In offline model-based RL settings, it is important to estimate the transition dynamics T^* with batch data from an expert to understand their behaviour



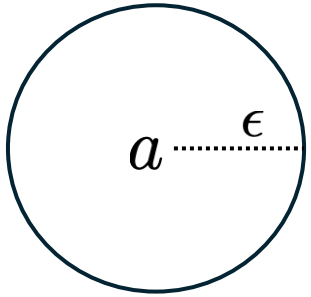
Goal: From batch data, determine the transition dynamics T^* that produce an ϵ-optimal expert policy under a reward model R

Inverse Transition Learning

We introduce a set of constraints to infer the true transition dynamics T^* when given access to an expert policy

Constraint 1

$$Q^*(s,a;T) - Q^*(s,a';T) \geq \epsilon$$
$$\iff$$
$$R(s,a) - R(s,a') + \gamma\big(T(.|s,a) - T(.|s,a')\big)^\top \times (I - \gamma T_{\pi^*(T)})^{-1} R_{\pi^*(T)} \geq \epsilon$$



Constraint 2

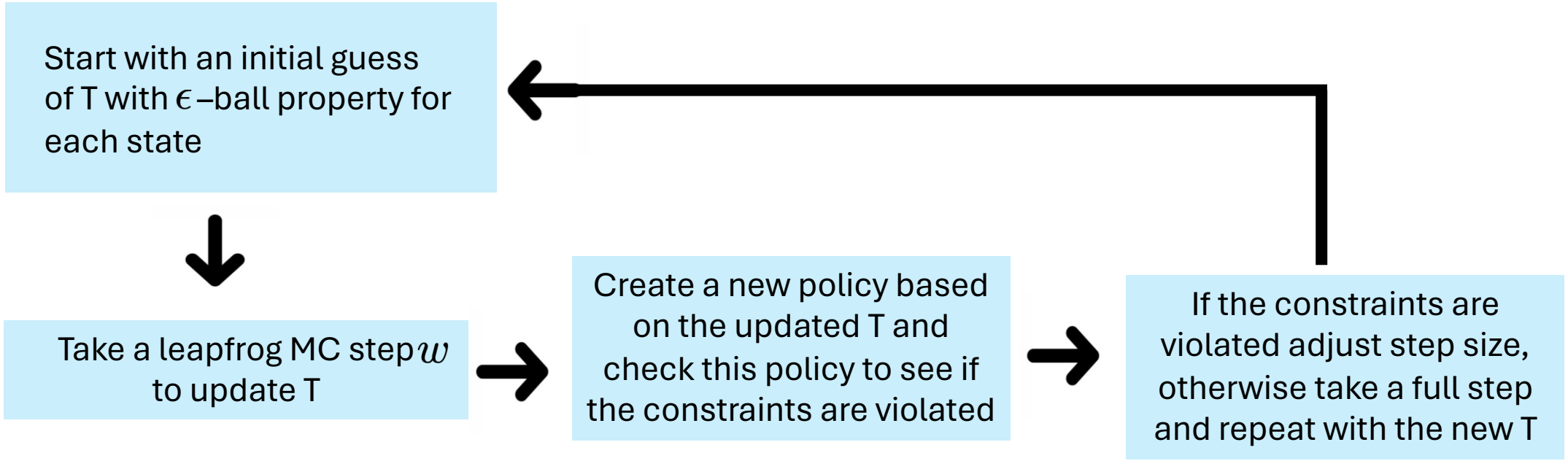
$$|Q^*(s,a;T) - Q^*(s,a';T)| \leq \epsilon$$
$$\iff$$
$$|R(s,a) - R(s,a') + \gamma\big(T(.|s,a) - T(.|s,a')\big)^\top \times (I - \gamma T_{\pi^*(T)})^{-1} R_{\pi^*(T)}| \leq \epsilon$$

Ensures that valid actions are at least ϵ away from invalid actions for every state

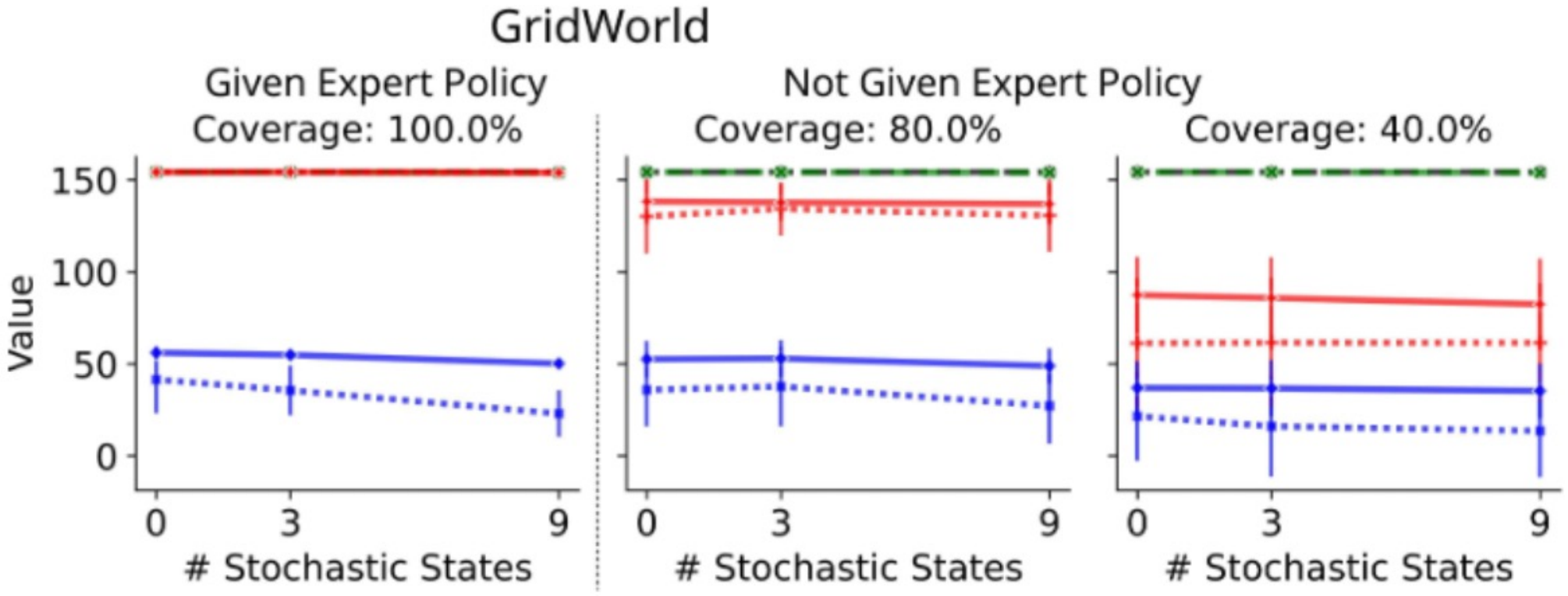
Ensures every valid action is close to every other valid action

Inverse Transition Learning

In the absence of an expert policy, we infer T^* from batch data alone. We estimate a posterior over transition dynamics using **Hamiltonian Monte Carlo with Reflection**.



Policies based on ITL outperform baselines, under varying degrees of coverage and optimality



On clinical tasks, ITL outperforms baselines in high-dimensional constrained settings

- Compare how often our method selects the action chosen by the expert in dataset and how often it selects a valid action within an ϵ-ball of the expert on MIMIC-IV task.

	Standard Task				Transfer Task			
	T^{MLE}	$\widehat{T}^*$	$P(T^*\|\mathcal{D})$	$P_\epsilon(T^*\|\mathcal{D})$	T^{MLE}	$\widehat{T}^*$	$P(T^*\|\mathcal{D})$	$P_\epsilon(T^*\|\mathcal{D})$
Best Action Matching over $\mathcal{D}$	0.348	**0.523**	0.348	**0.465**	0.372	**0.51**	0.37	**0.477**
ϵ-Ball Action Matching over $\mathcal{D}$	0.589	**1.00**	0.589	**1.00**	0.678	**0.966**	0.629	**0.97**

- Results do not change for varying degrees of optimality of expert

ITL enables us to reason about counterfactuals effectively

A



Normal Blood Pressure

Poor Consciouness

Good Blood Oxygen

Elevated Creatinine

B



Severe Hypotension

Ave Consciouness

Poor Blood Oxygen

Elevated Creatinine



Expert Policy

A

Prescribe **bolus** instead of vasopressor and bolus → Full recovery after 72 hrs

B

Prescribe **both vasopressor and bolus** instead of just a bolus → Relocated after 72 hrs to general ward

Summary

- ITL is a new approach for offline model-based RL for accurate estimation the transition dynamics by leveraging expert demonstrations.

- Our approach outperforms traditional model-based RL methods in both high-dimensional and constrained environments.

- ITL enables us to extract meaningful insights and improve decision-making processes in tasks such as managing patients in ICU.

Please reach out to us on slack or email s.parbhoo@imperial.ac.uk for more details