Confidence Aware Inverse Constrained Reinforcement Learning

Sriram Ganapathi Subramanian[1], Guiliang Liu[2], Mohammed Elmahgiubi[3], Kasra Rezaee[3], Pascal Poupart[1,4]

[1]Vector Institute,
[2]Chinese University of Hong Kong, Shenzhen,
[3]Huawei,
[4]University of Waterloo

Presentation at ICML 2024

VECTOR INSTITUTE

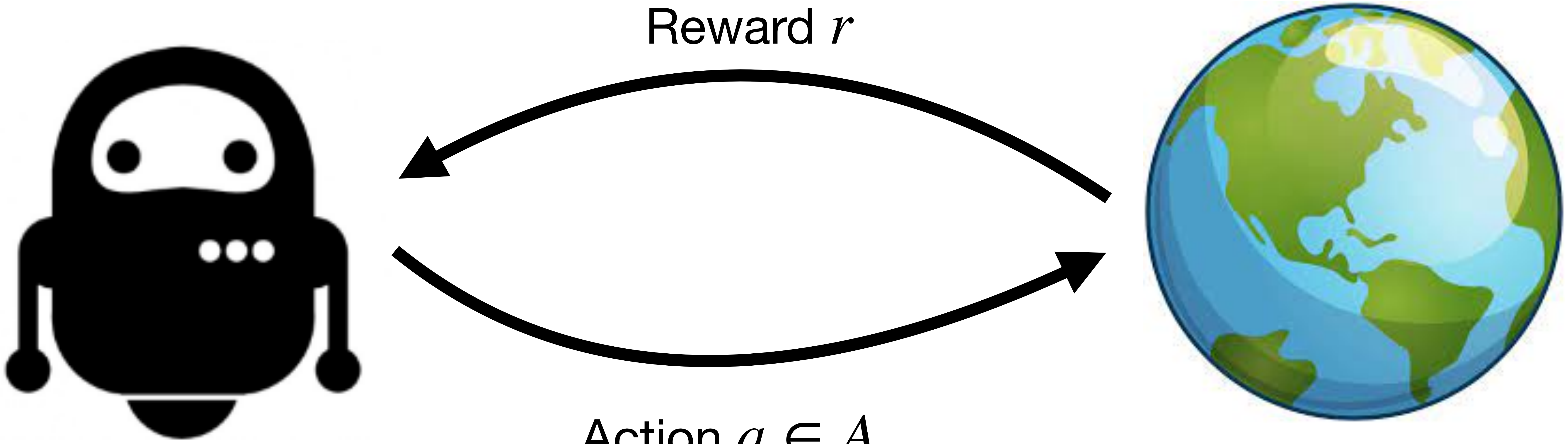
香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

HUAWEI

UNIVERSITY OF WATERLOO

Reinforcement Learning

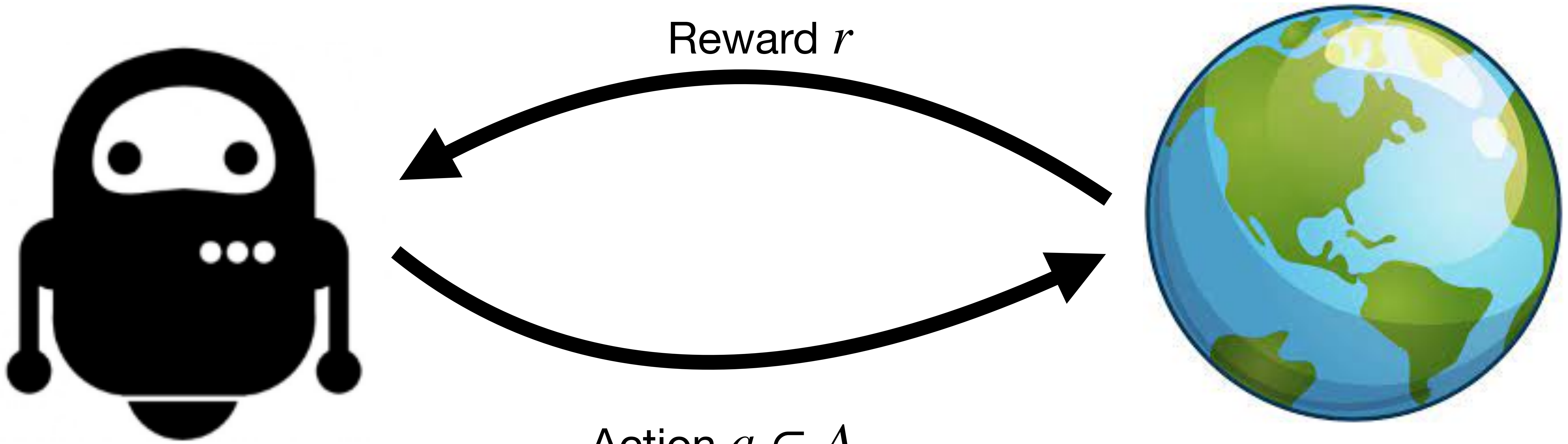
Objective: Find the optimal policy that maximizes the expected discounted cumulative rewards



Best policy maximizes the total expected return: $V_\pi = \mathbb{E}_\pi \sum_t \gamma^t r_t$

Constrained Reinforcement Learning (CRL)

Objective: Find the optimal policy that maximizes the expected discounted cumulative rewards subject to constraints

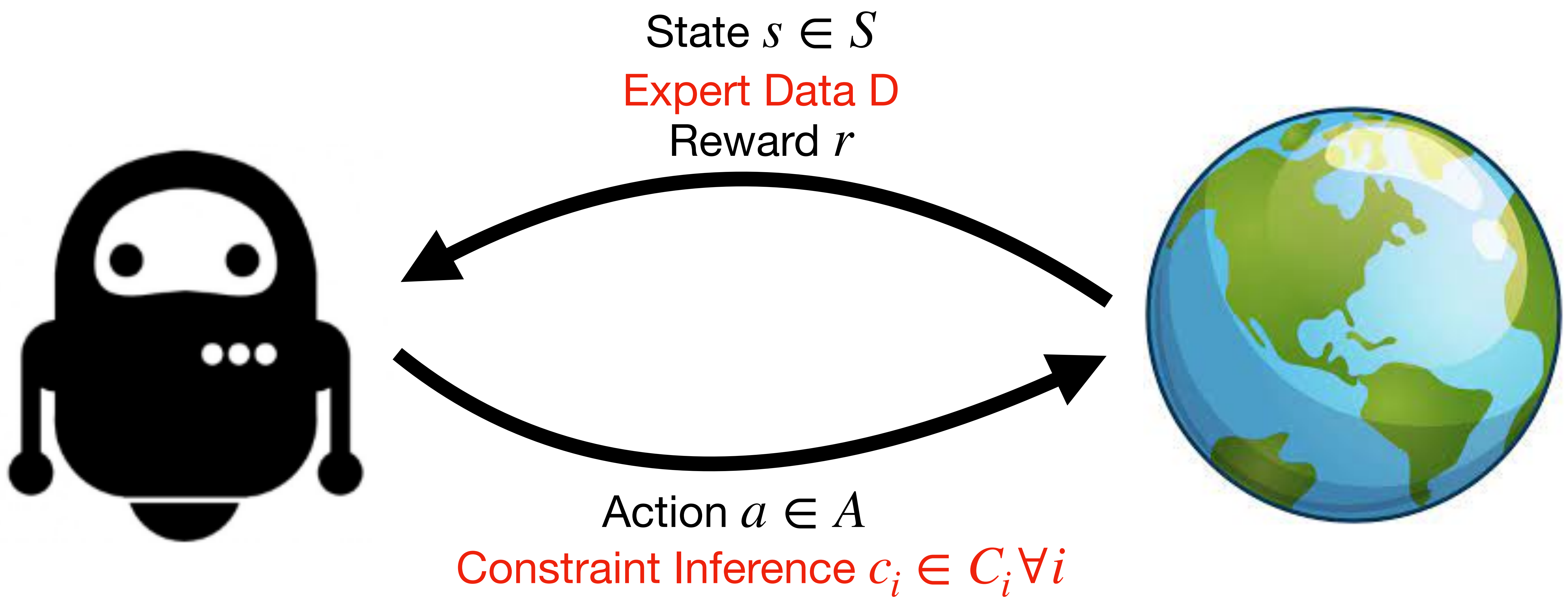


Best policy maximizes the total expected return: $V_\pi = \mathbb{E}_\pi \sum_t \gamma^t r_t$

such that $\mathbb{E}_\pi [\sum_t \gamma^t c_i(s_t, a_t)] \leq \beta_i \forall i$

Inverse Constrained Reinforcement Learning

Objective: Given a dataset of expert demonstrations and (optionally) a reward function, find constraints (and optional rewards) such that when CRL is performed, we obtain a constrained optimal policy that represents the behaviour in the dataset



Agent performs constraint inference using expert data and learns a constrained policy model using the estimated constraints at the same time

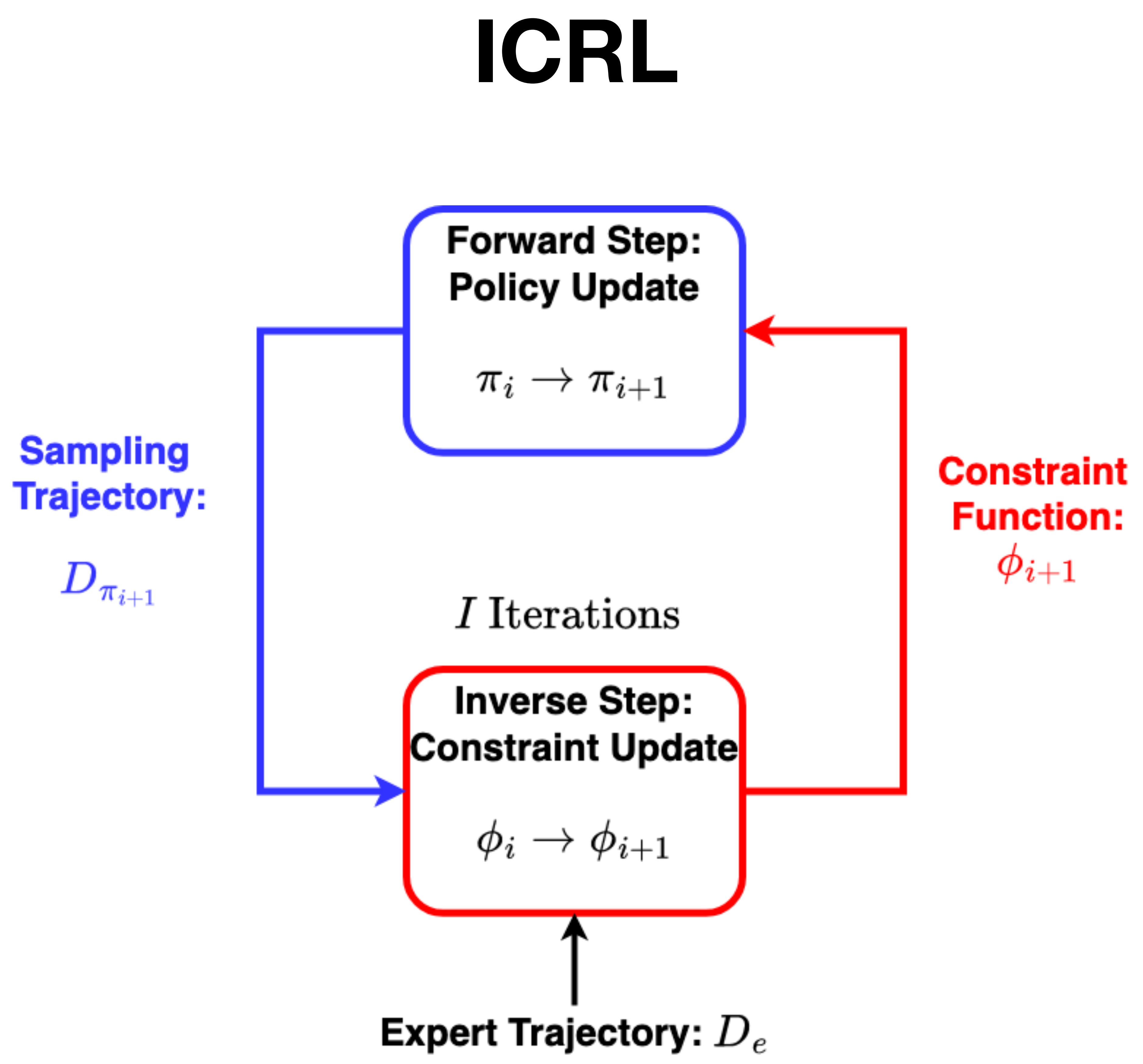
Confidence Aware Inverse Constrained RL

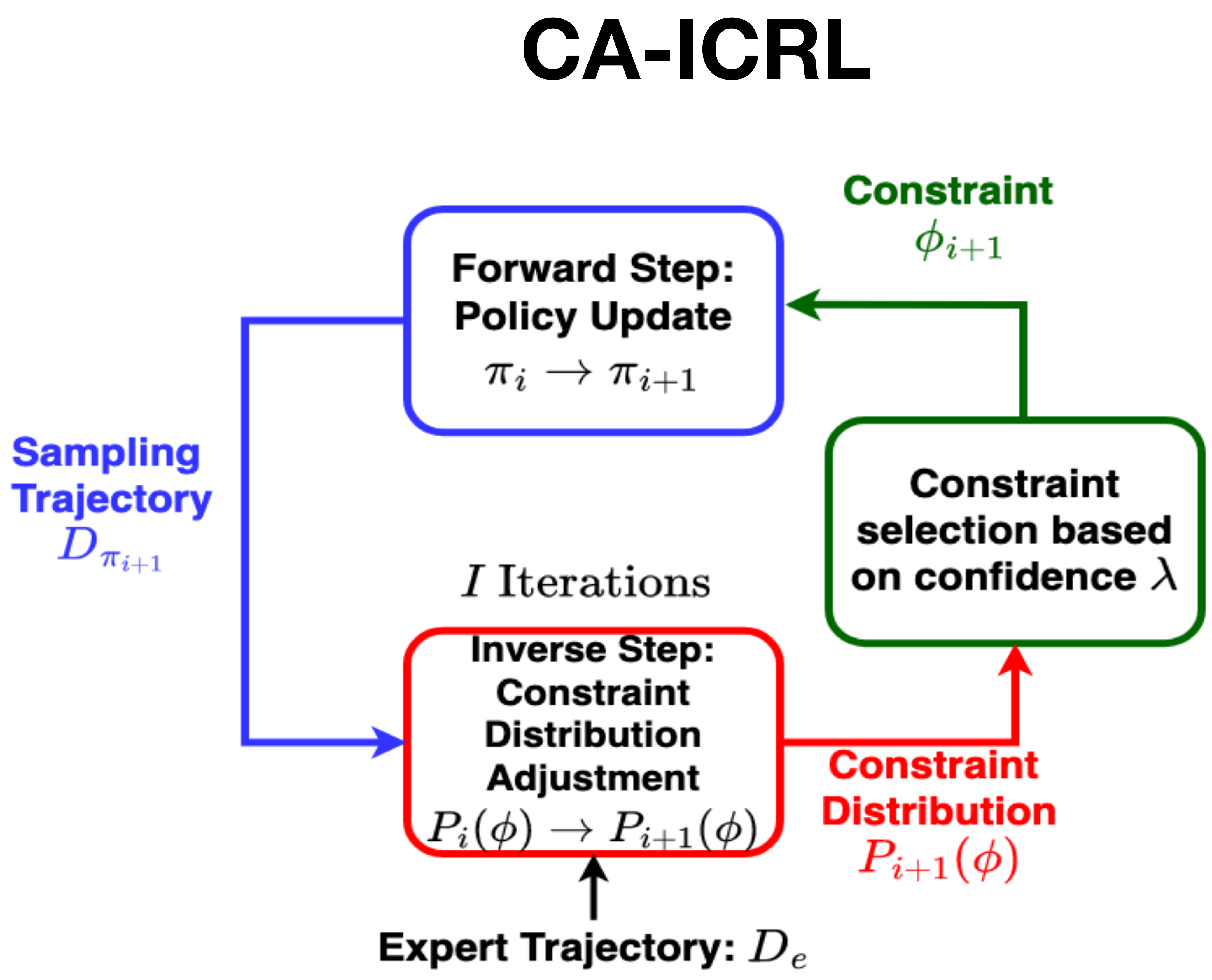
- Intuitively more expert trajectories adhering to a constraint implies higher confidence

- CA-ICRL maintains a confidence estimate along with learning constraints

- Two use cases:
  - Inferring constraints based on confidence level
  - Determining sufficiency of expert trajectories based on desired confidence and performance

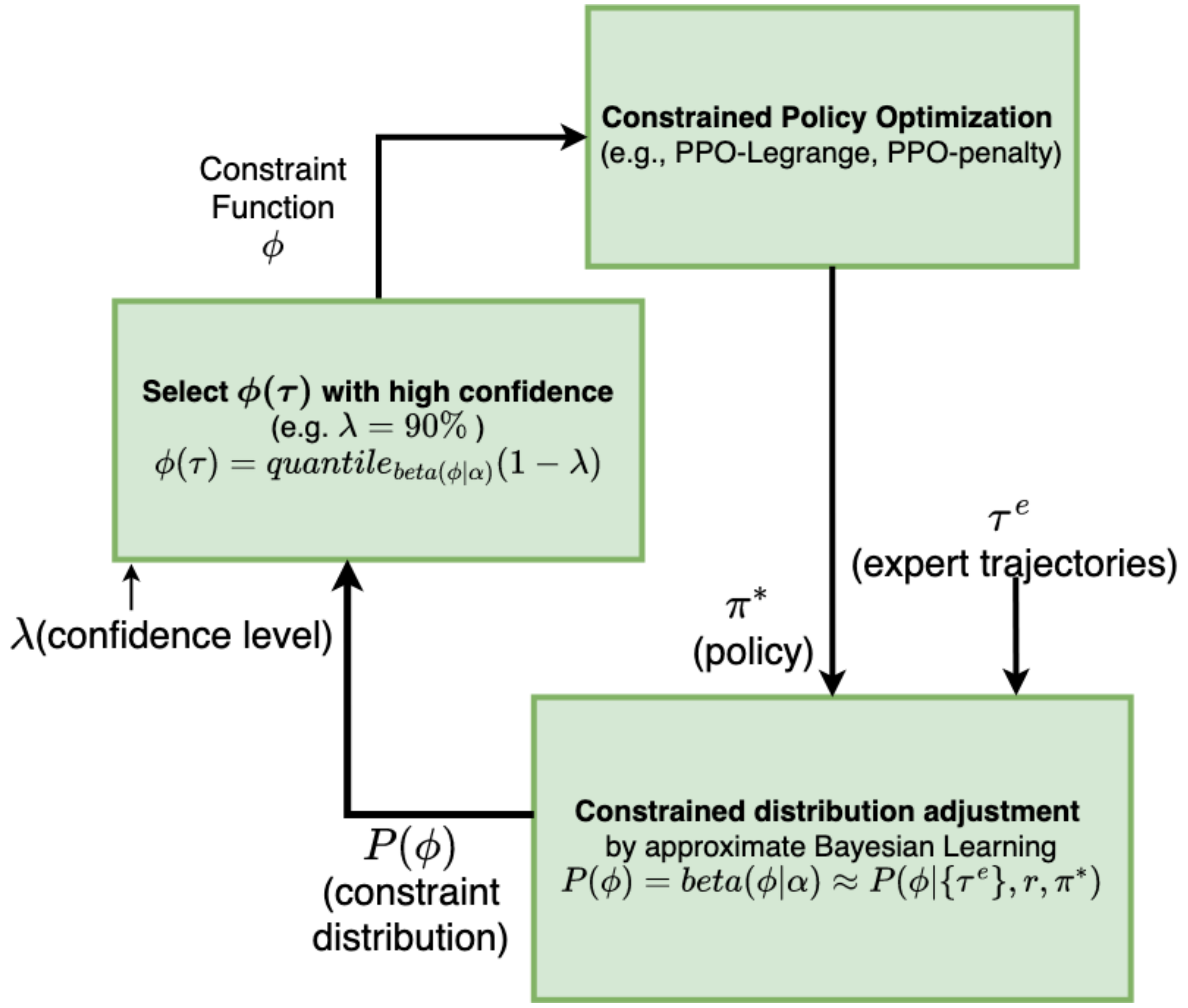
ICRL vs CA-ICRL



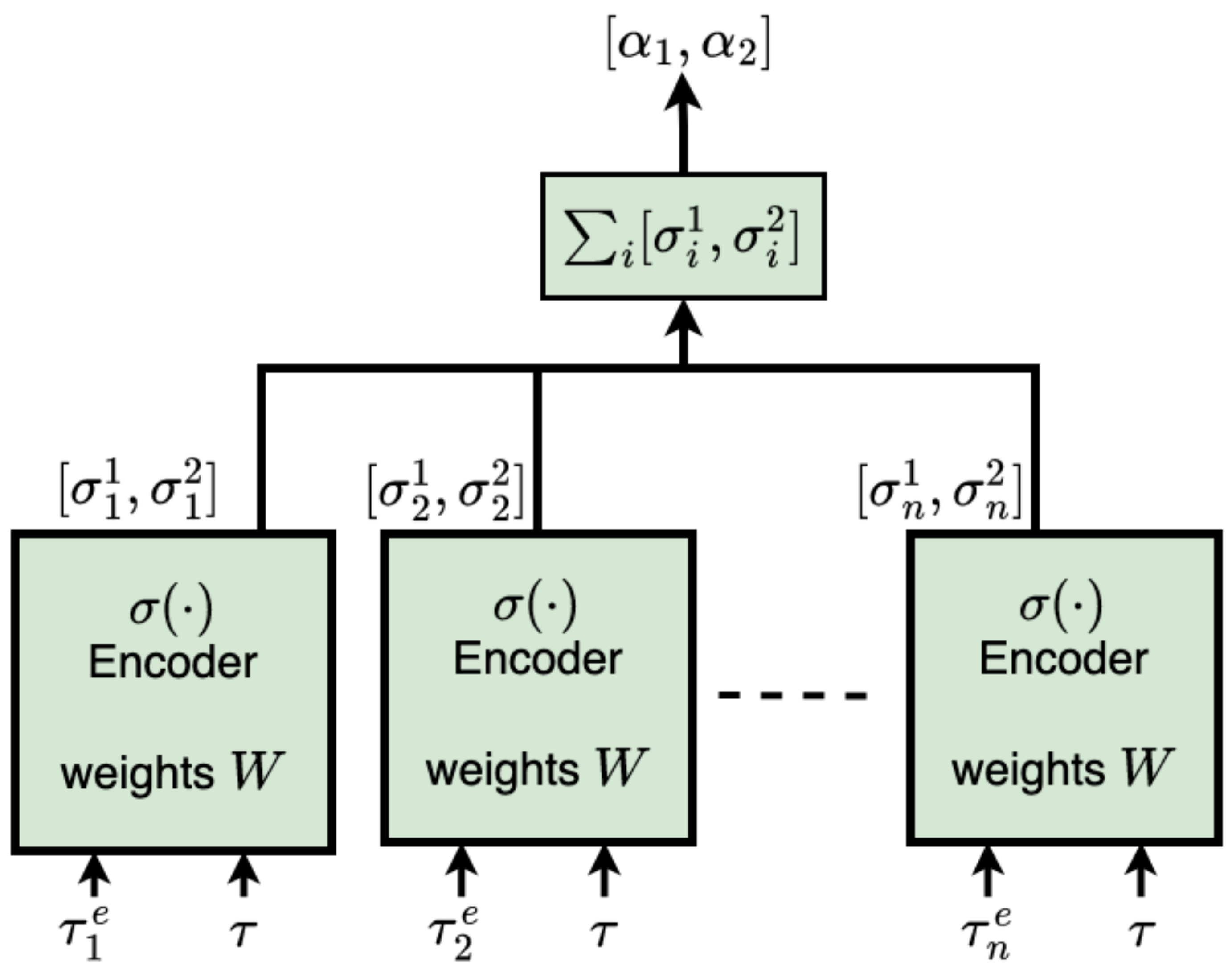


Case 1: Inferring a constraint conditioned on a confidence level

- Objective: Inferring constraints with the provided (desired) confidence
- A beta distribution is used to model the constraint function
- Distribution is represented by a neural network having multiple encoder blocks with shared weights



Confidence Aware Constraint Inference

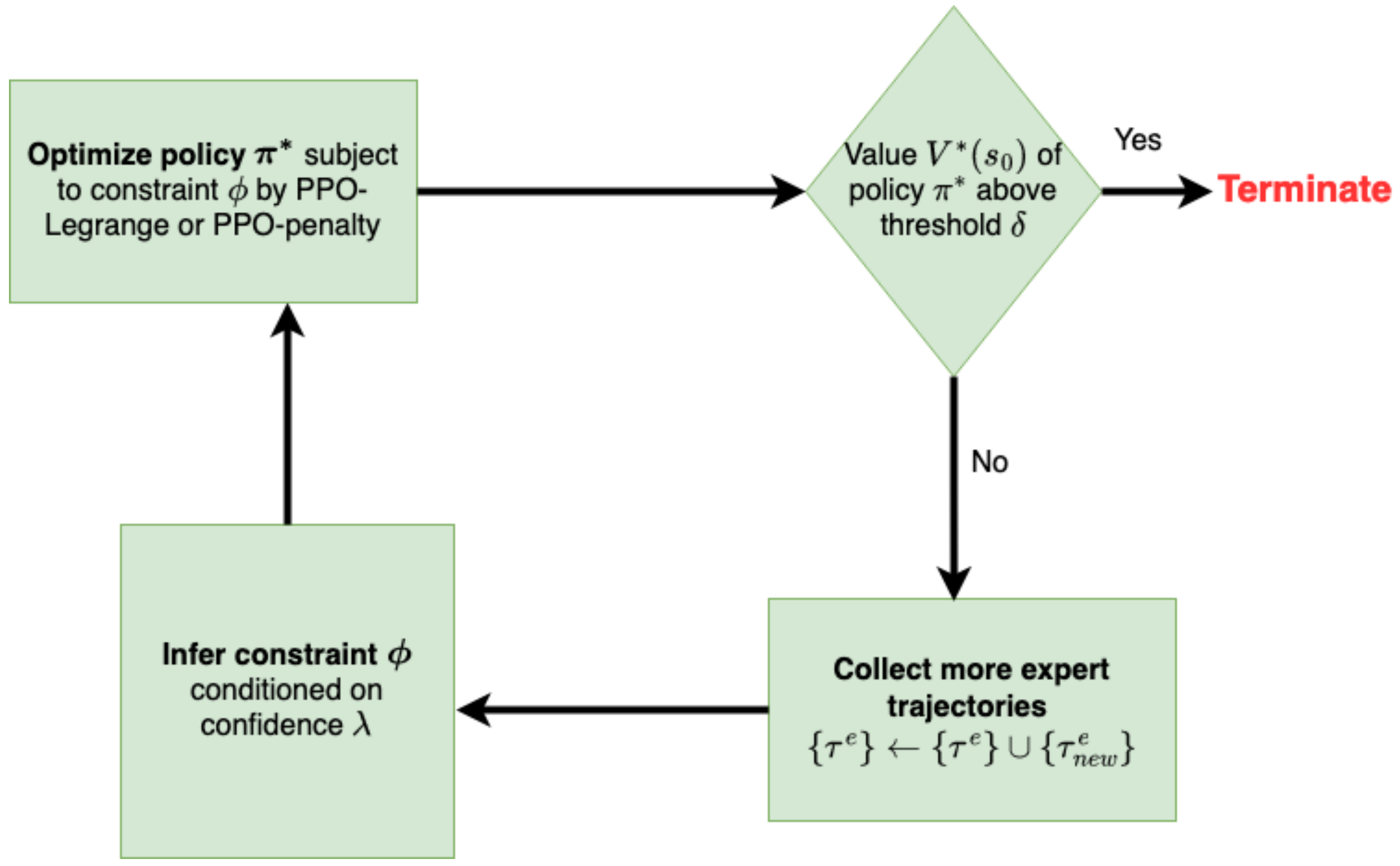


Network Architecture for Constraints with Confidence

Case 2: Determining sufficiency of expert trajectories

- Objective: Determining if more expert trajectories are required based on the desired performance

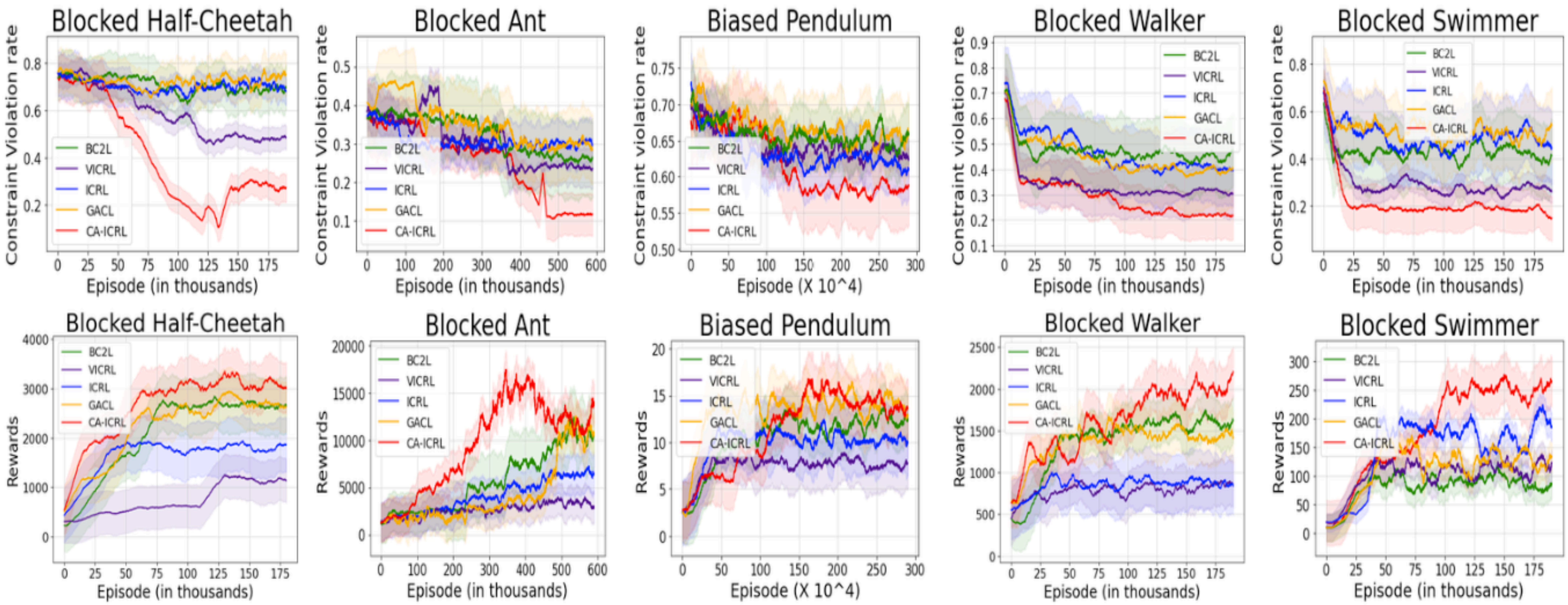
- Less expert trajectories $\Longrightarrow$ more constraining learned constraints $\Longrightarrow$ less performance by learned policies



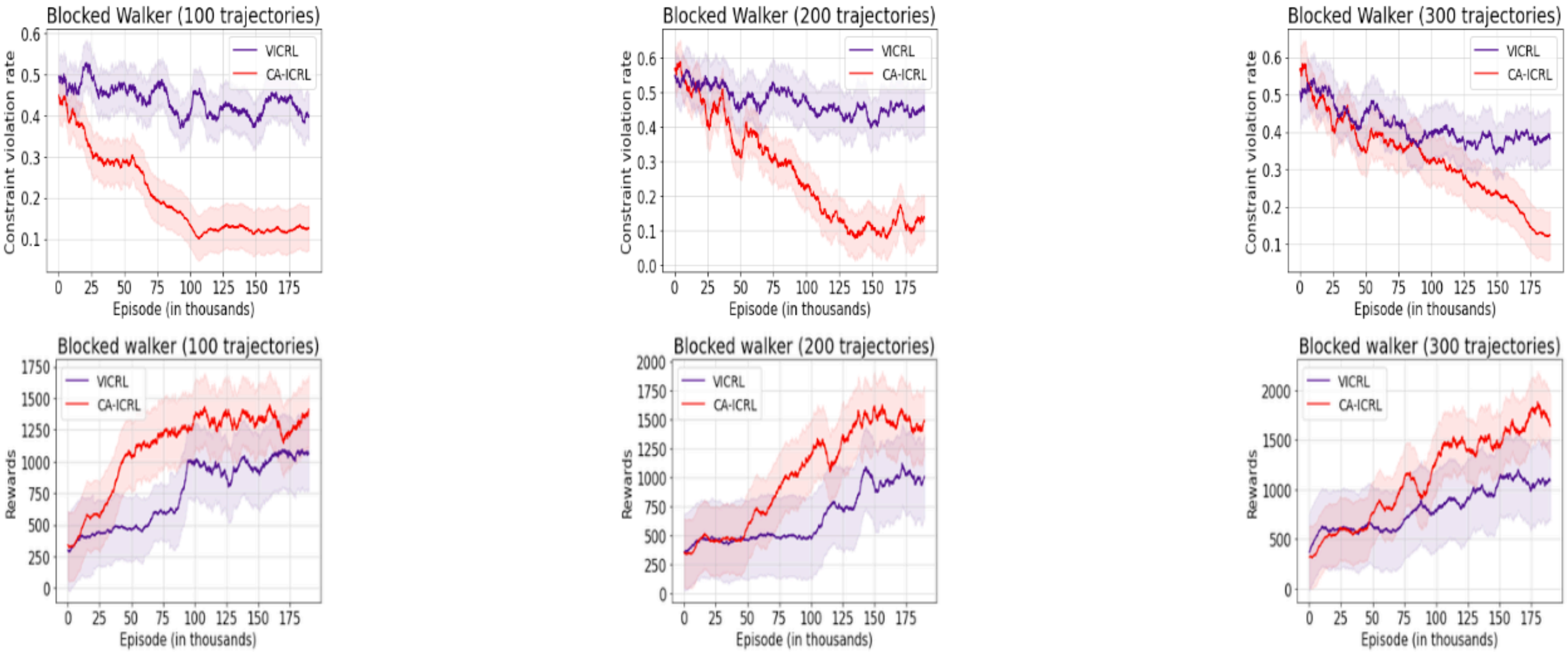
Experiments

- Variety of experiments covering multiple domains
  - 5 robotics domains as modifications of well-known MuJoCo environments
  - Two real world domains where data comes from realistic HighD dataset
- Extensive comparisons to several state of the art baselines: BC2L, GACL, ICRL, VICRL
- Performance measures: Constraint violation rate and reward obtained (need lower constraint violation rate and higher rewards)

Results: Case 1 ($\lambda = 0.7$)



Results: Case 2 ($\lambda = 0.8$)



Conclusion

- Introduced a notion of confidence in ICRL

- Two use cases:
    - Inferring constraints with desired confidence
    - Determining sufficiency of expert trajectories with desired confidence and performance

- Extensive experimental results on robotics and driving domains
    - Better performances than baselines
    - Learned constraints at-least as constraining as ground truth constraints with desired confidence

Thank you

Full paper - https://arxiv.org/pdf/2406.16782

Source Code - https://github.com/Sriram94/ConfidenceAwareICRL

Questions or Comments?

Please send an email to Sriram Ganapathi Subramanian: sriram.subramanian@vectorinstitute.ai

Personal Website: https://sriramsubramanian.com/

VECTOR INSTITUTE

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

HUAWEI

UNIVERSITY OF WATERLOO