

EAGER: Hierarchical Contrastive Explanations for Robot-Human Communication

PI: Siddharth Srivastava, Arizona State University

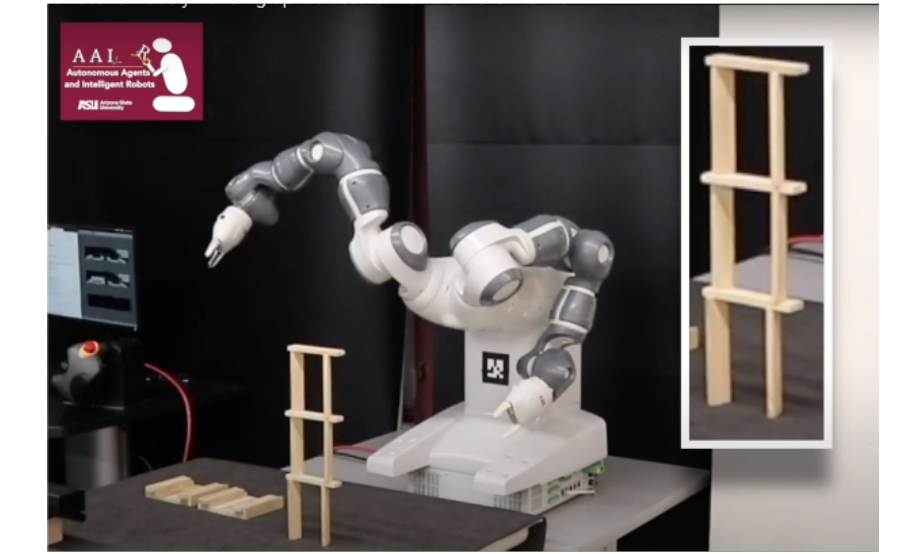
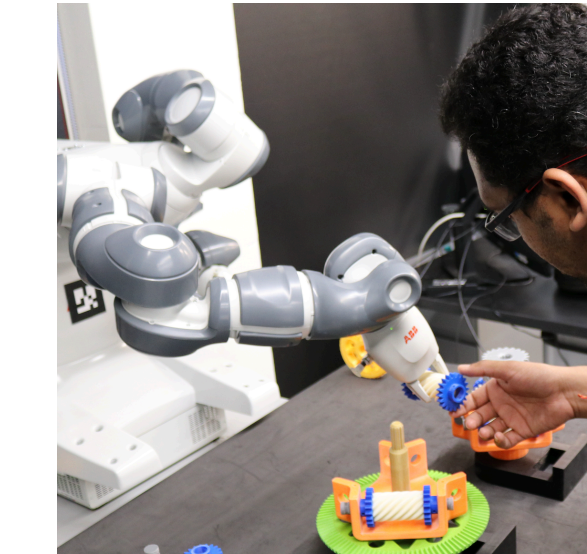
Co-PI: Subbarao Kambhampati, Arizona State University



How would a non-AI/robotics expert determine what their robot can and can't do?

Understand what it's doing and why?

Reconfigure it for a desired objective?



Key Challenges

1. User needs to be able to ask the **right questions** to assess robot's capability for new tasks.
2. Robot needs to be able to **explain** itself. Explanations need to minimize the **computational cost of processing information**.

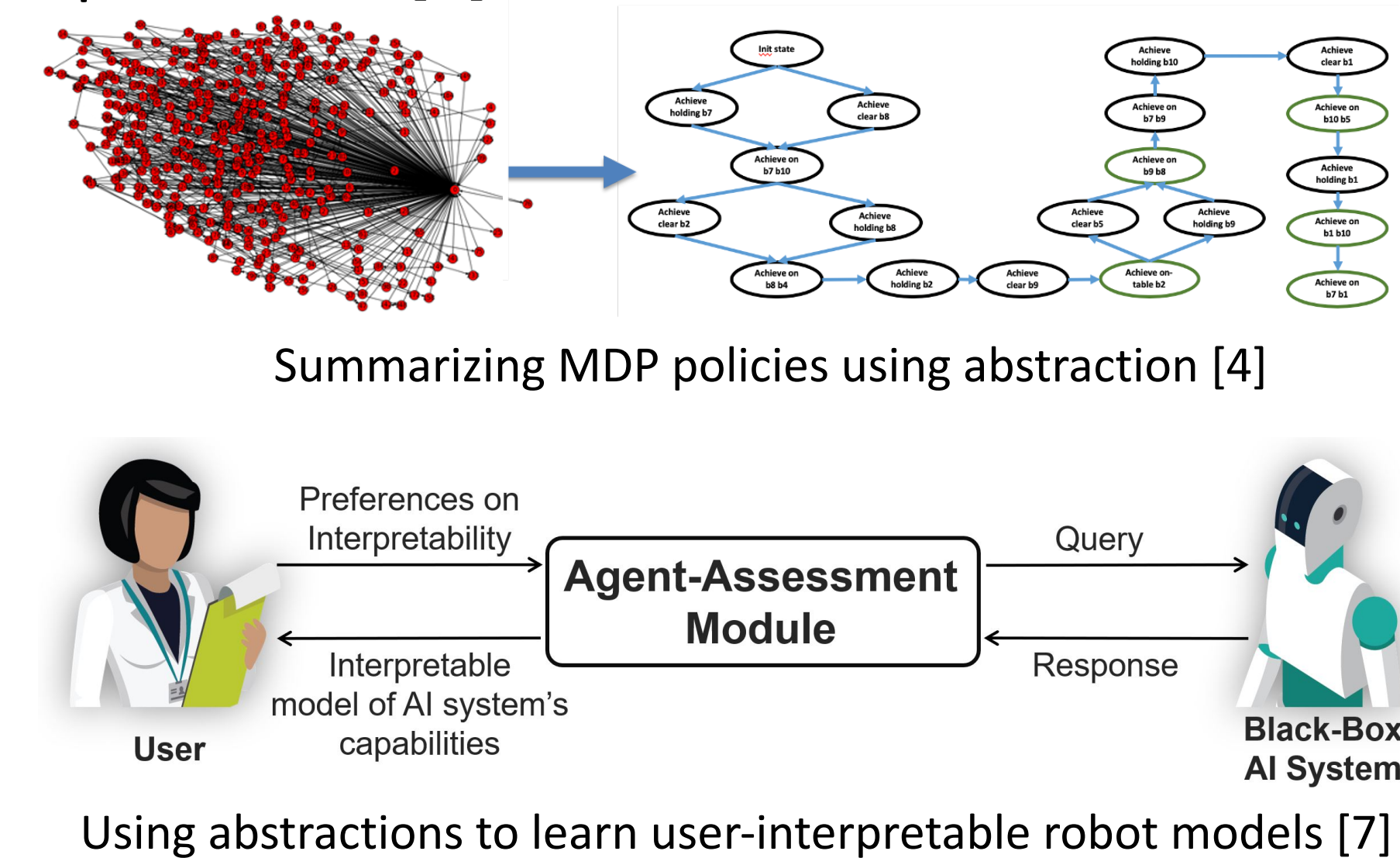
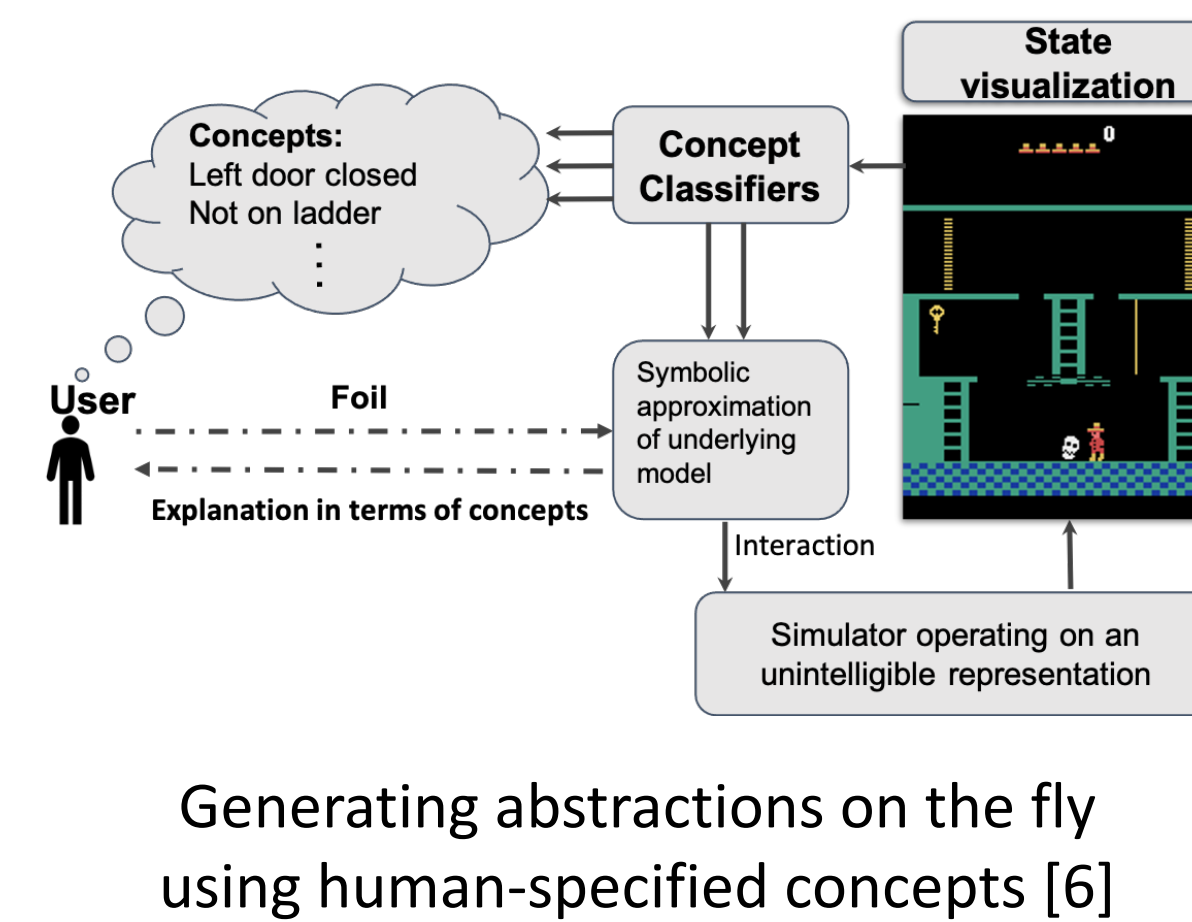
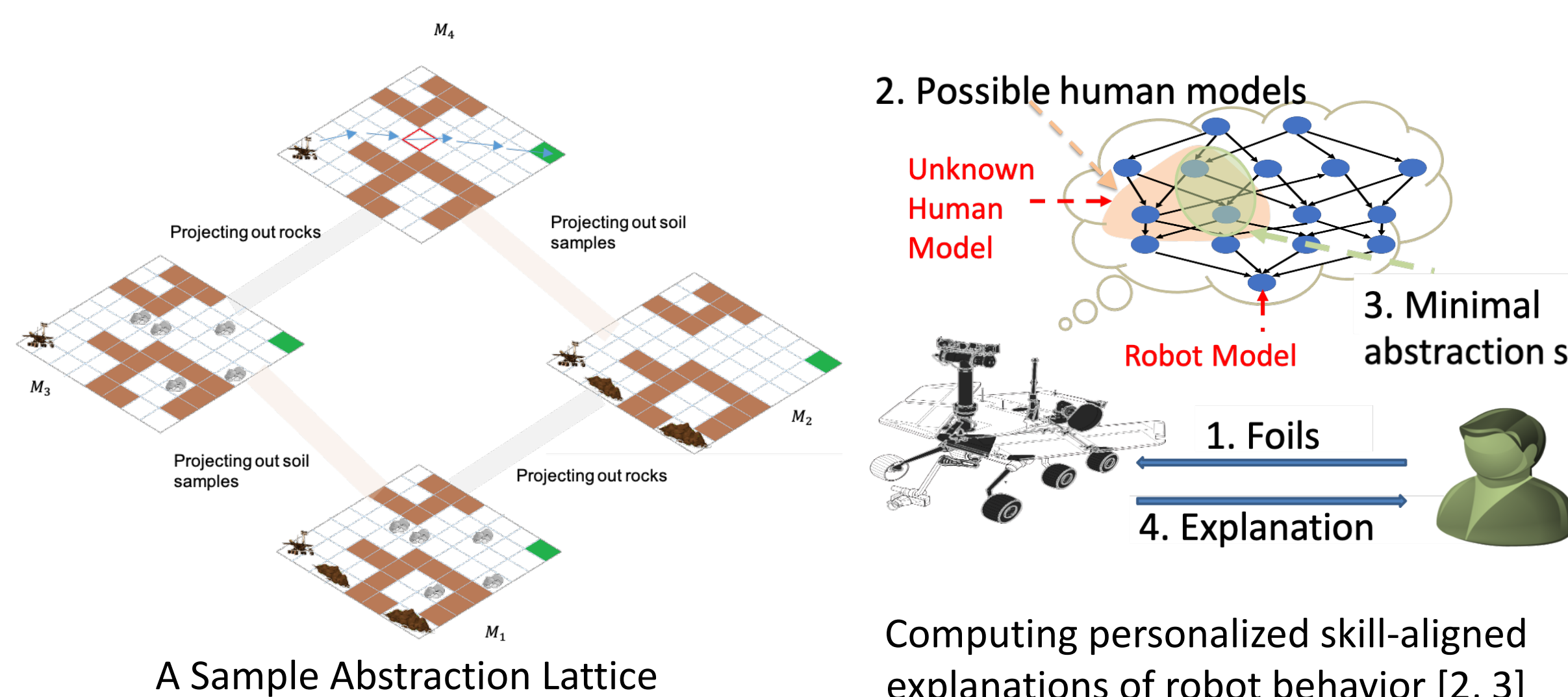
Scientific Impact

- Hierarchical explanations can be used to advance personalized training systems
- Query-based model estimation sheds new light on computing **interpretable models** for black-box, non-stationary cyber-physical systems and presents new algorithms for user-specific, personalized assessment of safe and assistive AI systems.

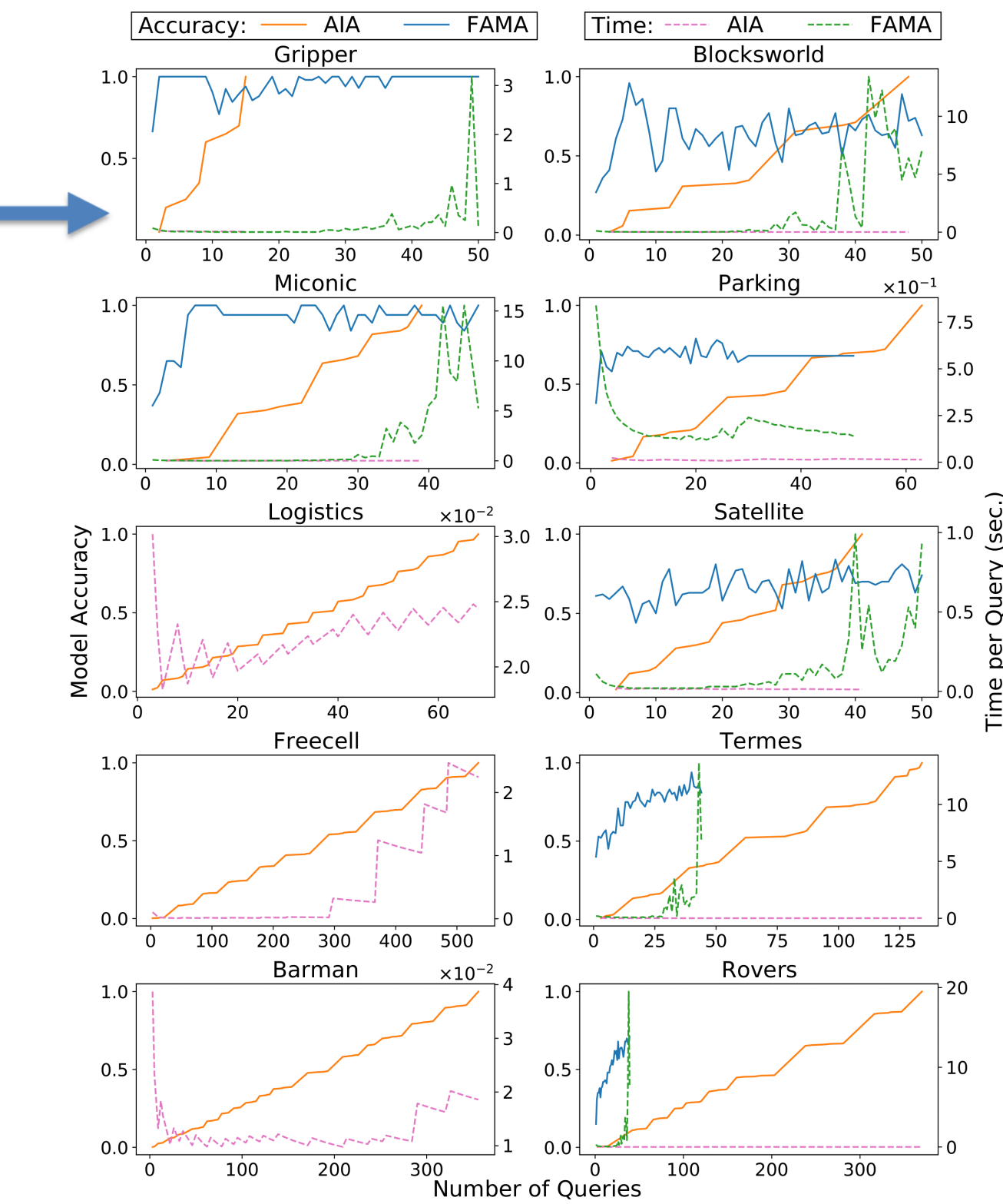
Solution Approach and Research Outcomes

Use new theory of **model abstractions** to define **lattice of abstract models** [1].

- A. Use lattice properties to provide explanations that minimize user's computational cost of processing information [2, 3, 7]
- B. Use lattice to compute hierarchical questioning strategy that constructs understandable model of robot's capabilities [4]
- C. Use an implicit lattice defined over user-specified concepts to identify model components for explanations [6]



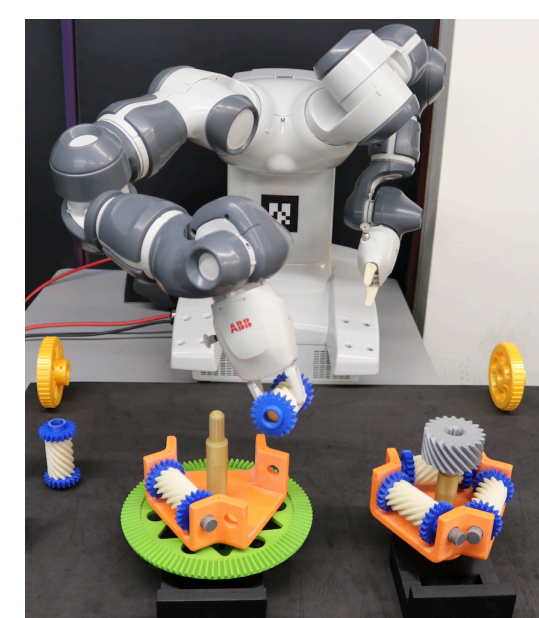
Our approach is faster, learns more accurate models than SoTA methods [7]



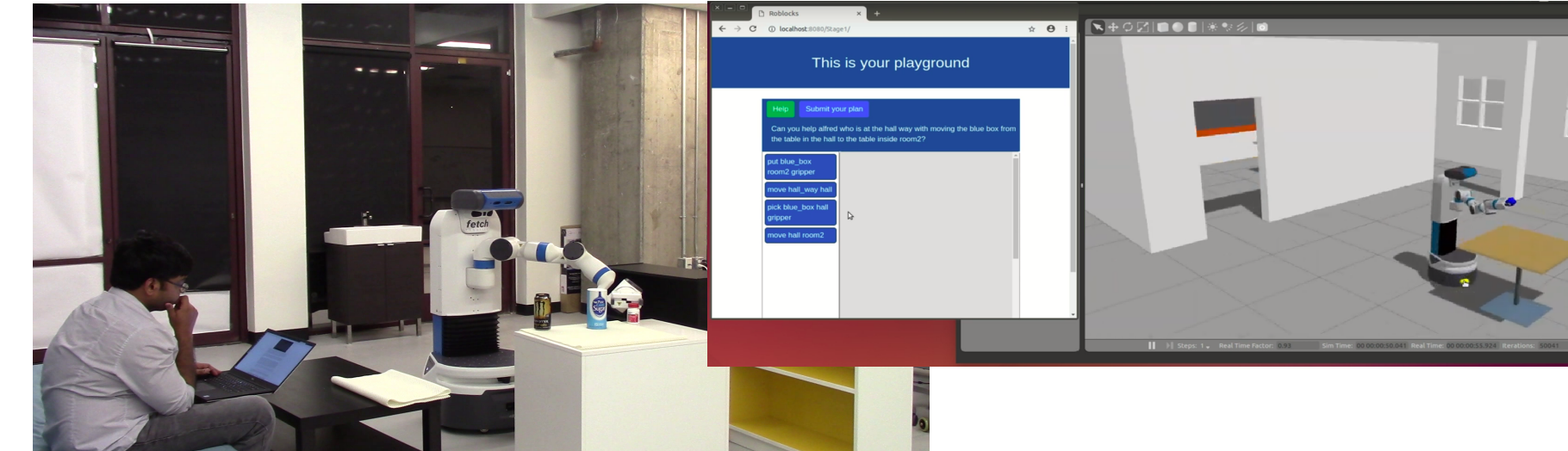
Broader Impact [Society]

Personalized, adaptive on-the-job training for working with AI robots

Increased worker employability and global competitiveness



Broader Impact [Education]



Adaptive AI/robotics education portals

Quantifiable Broader Impact

- ~2% (10%) of the US workforce has a bachelors' degree in computer science/mathematics (all science/engg)*
- This research develops foundations for enabling 90% of the workforce to use robots safely and effectively.

* NSF Science & Engineering Indicators 2018