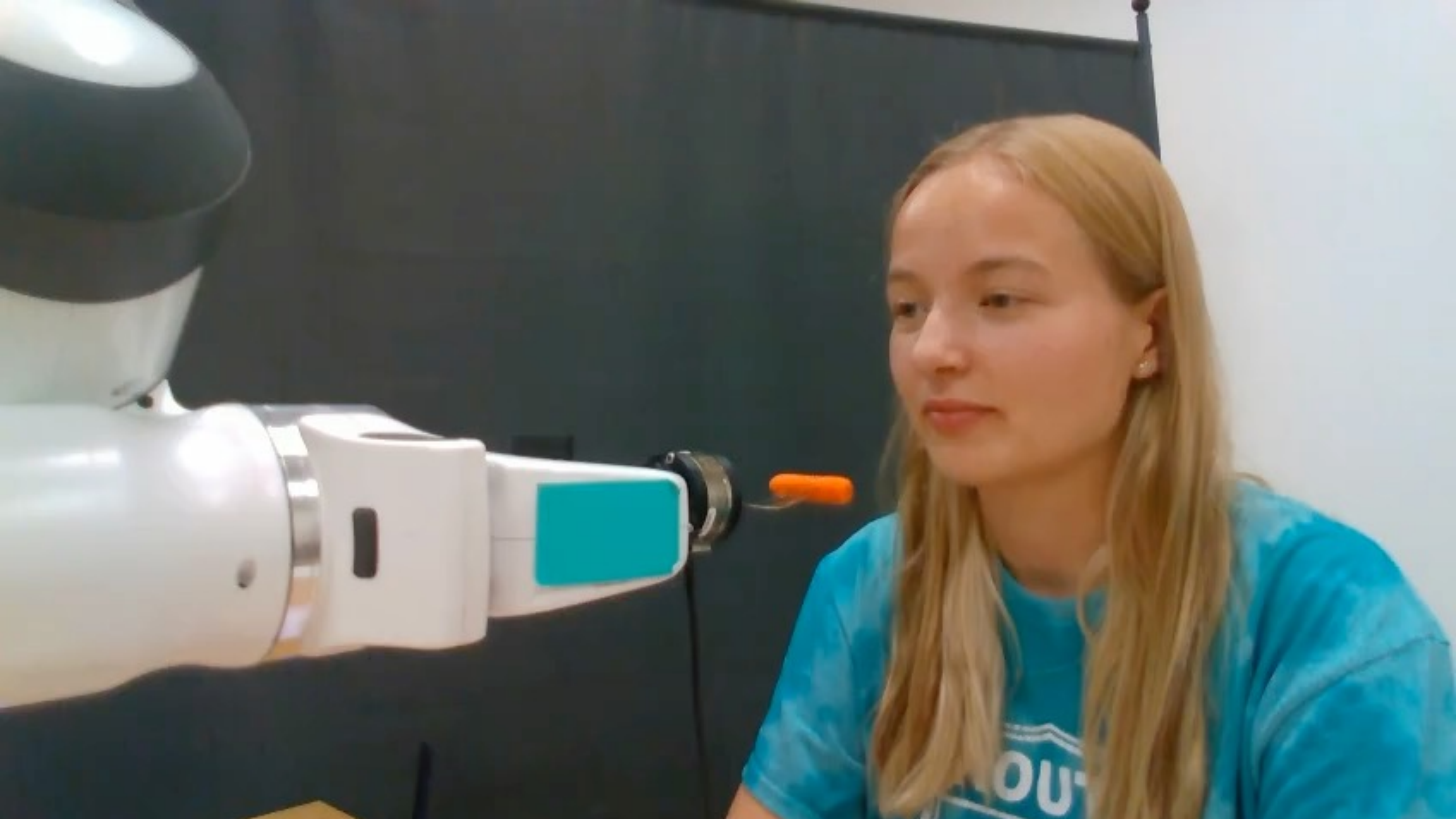
Interactive Learning in the Era of Large Models

Dorsa Sadigh

Stanford University

iliad
intelligent and interactive autonomous systems




Relying on limited expert demonstrations or reward signals is impractical!

Expert demonstrations are difficult to collect, variable, and suboptimal!

difficult to collect

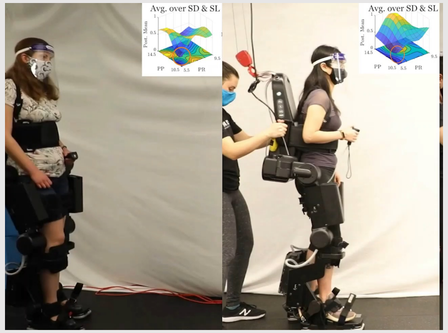
suboptimal and variable

[Basu et al. HRI 17]
[Kwon et al. HRI 20]


Demonstrations

Pairwise Comparisons **Physical Corrections** **Suboptimal Demonstrations**

Learning Human Preferences



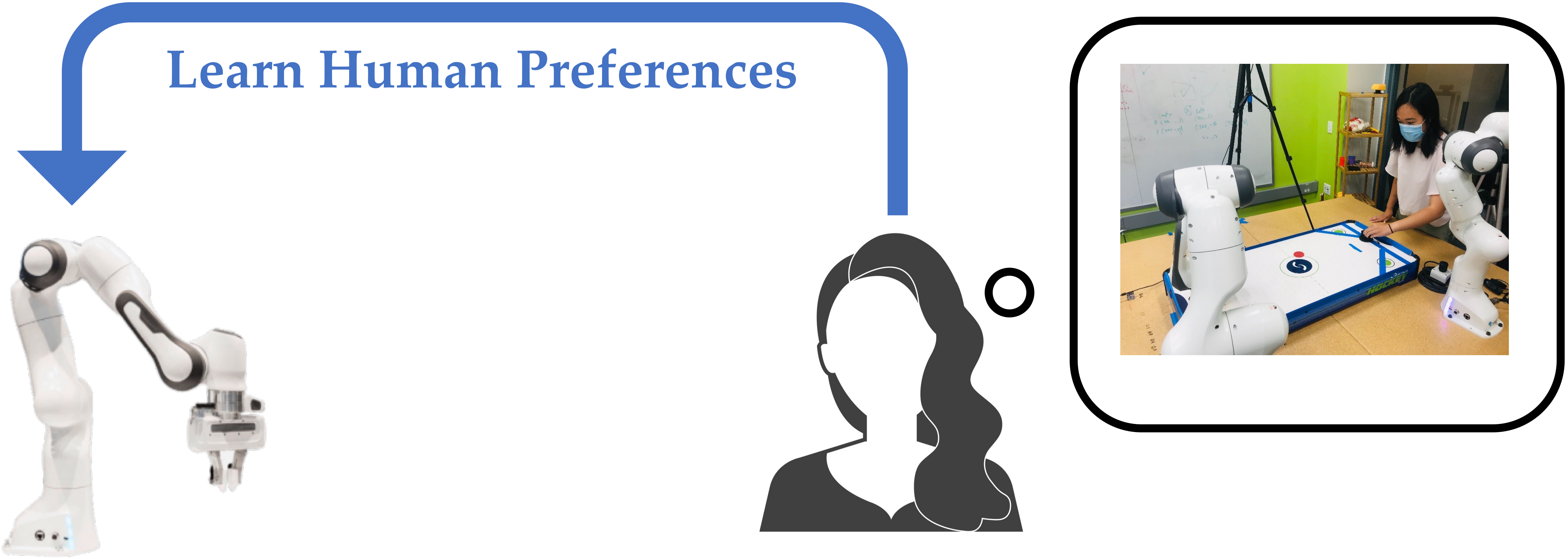
Biyik et al. IJRR 21
Kwon et al. ICLR 23
Gandhi et al. CoRL 22

Foundation Models for Robotics

Voltron

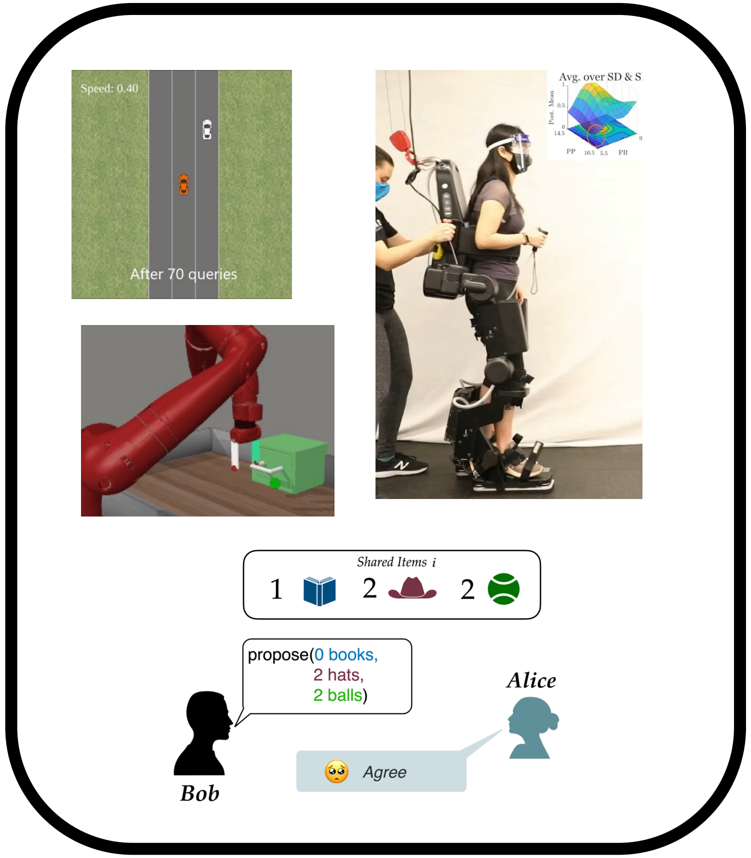
Karamcheti et al. RSS23
Mirchandani et al. CoRL23

How the human acts,

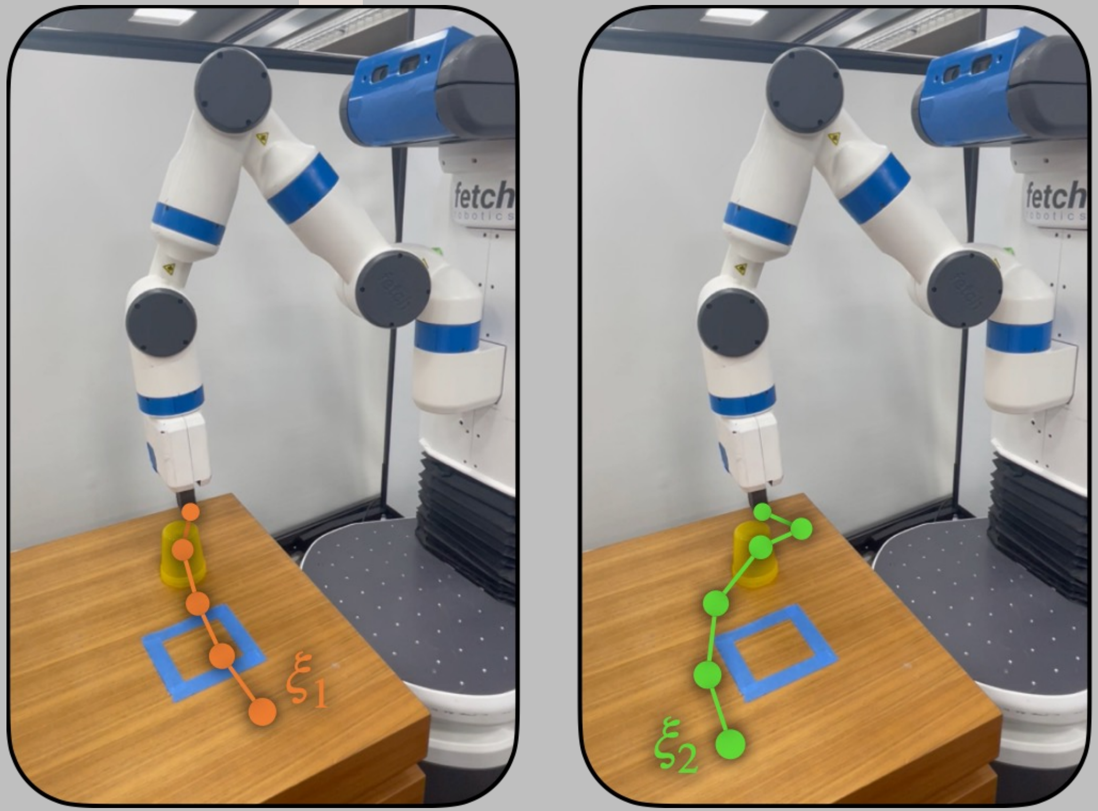


How the human acts,
but also how the human wants the robot to act

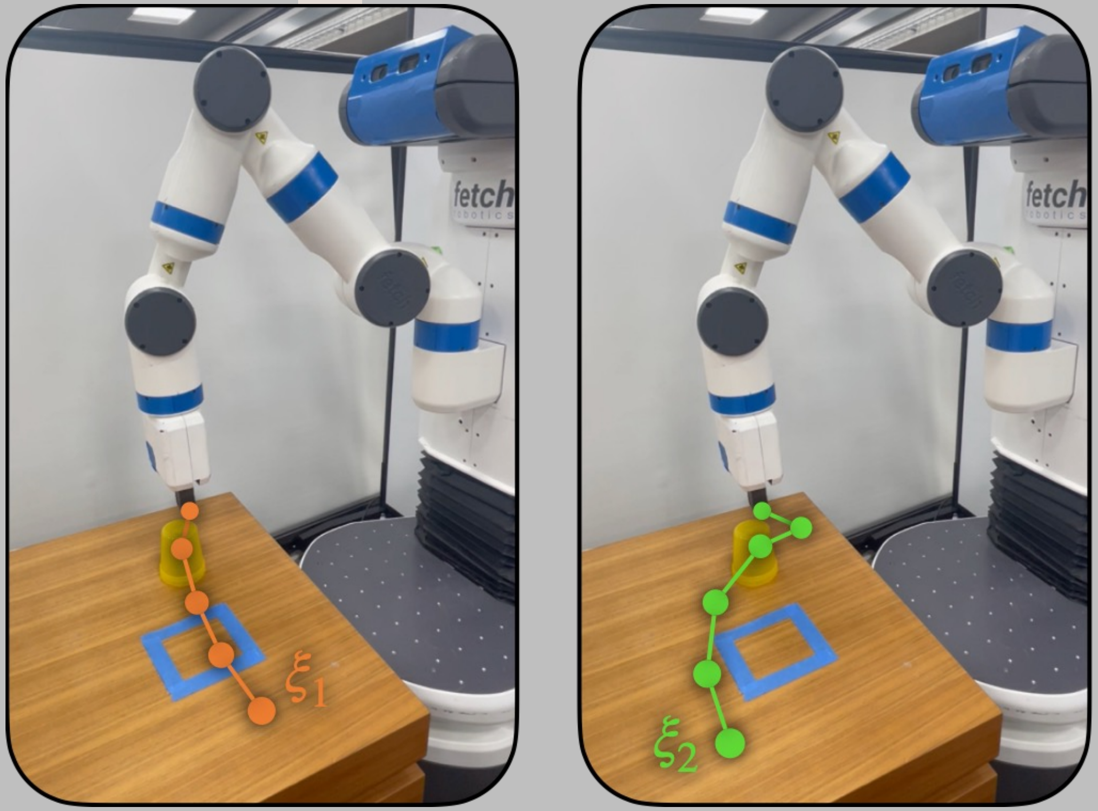
Learn Human Preferences



We need to learn representations of human preferences → Reward

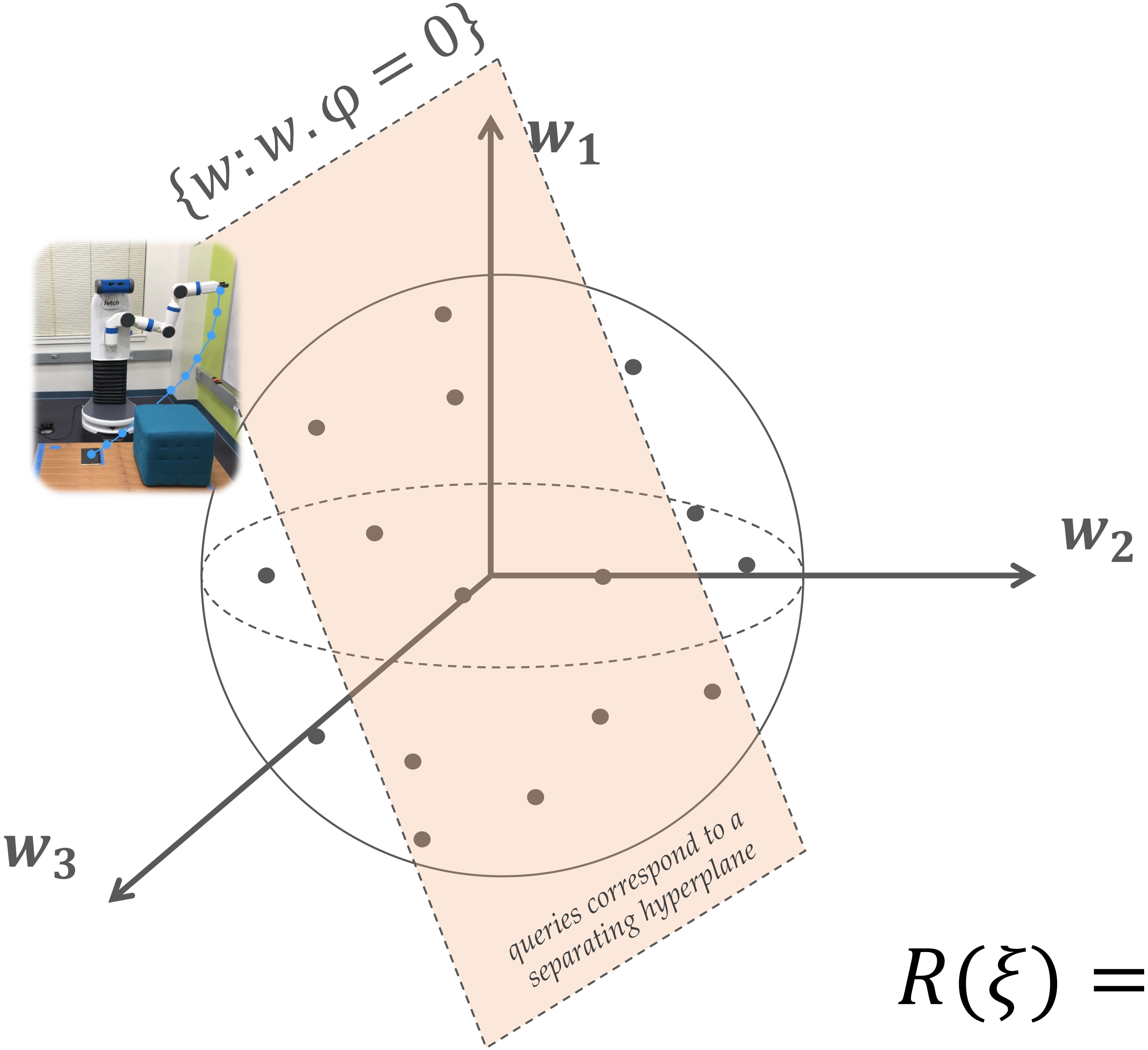


ξ_A or ξ_B?

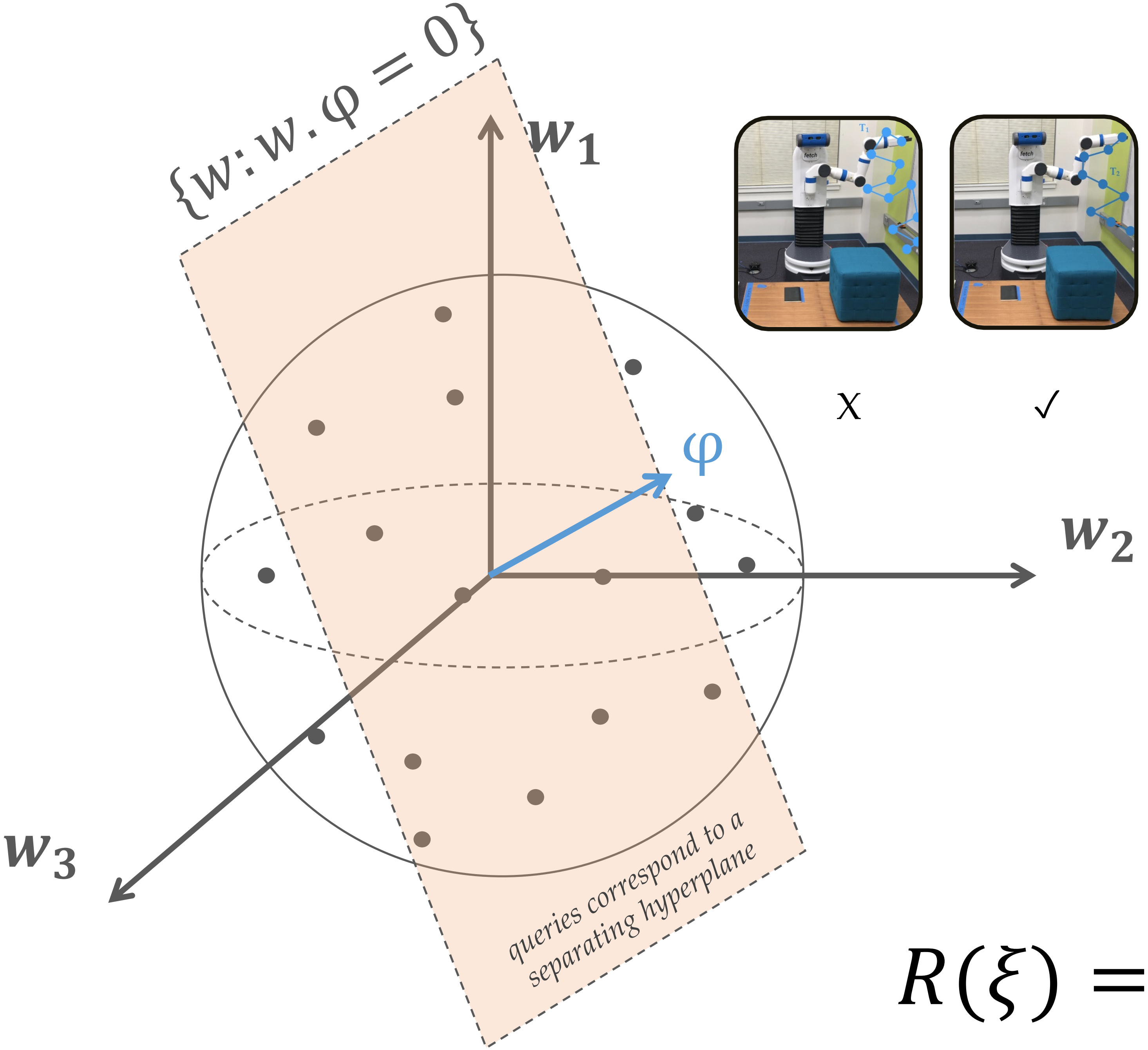


R_A or R_B?

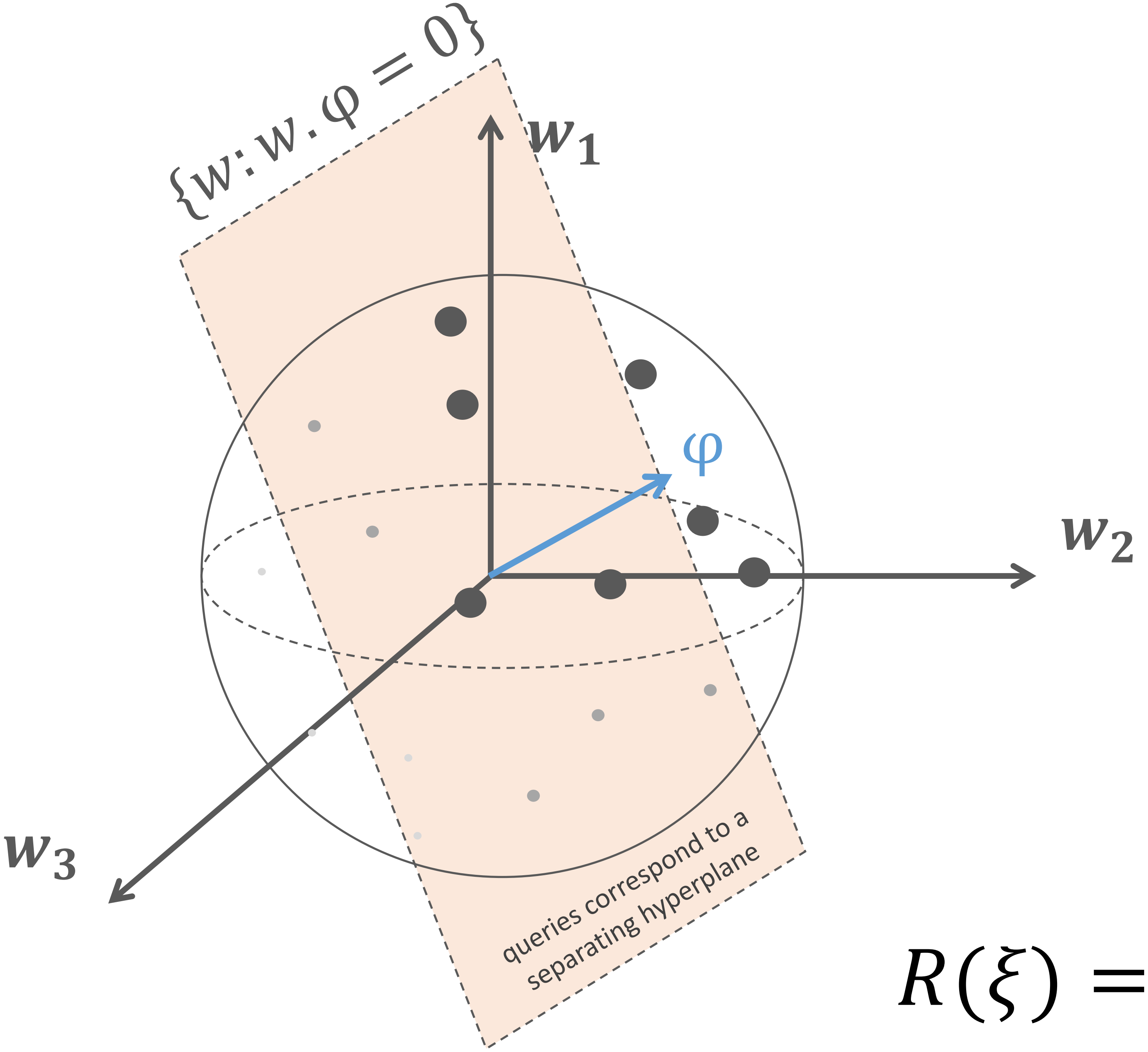
$$R(\xi) = w \cdot \phi(\xi)$$



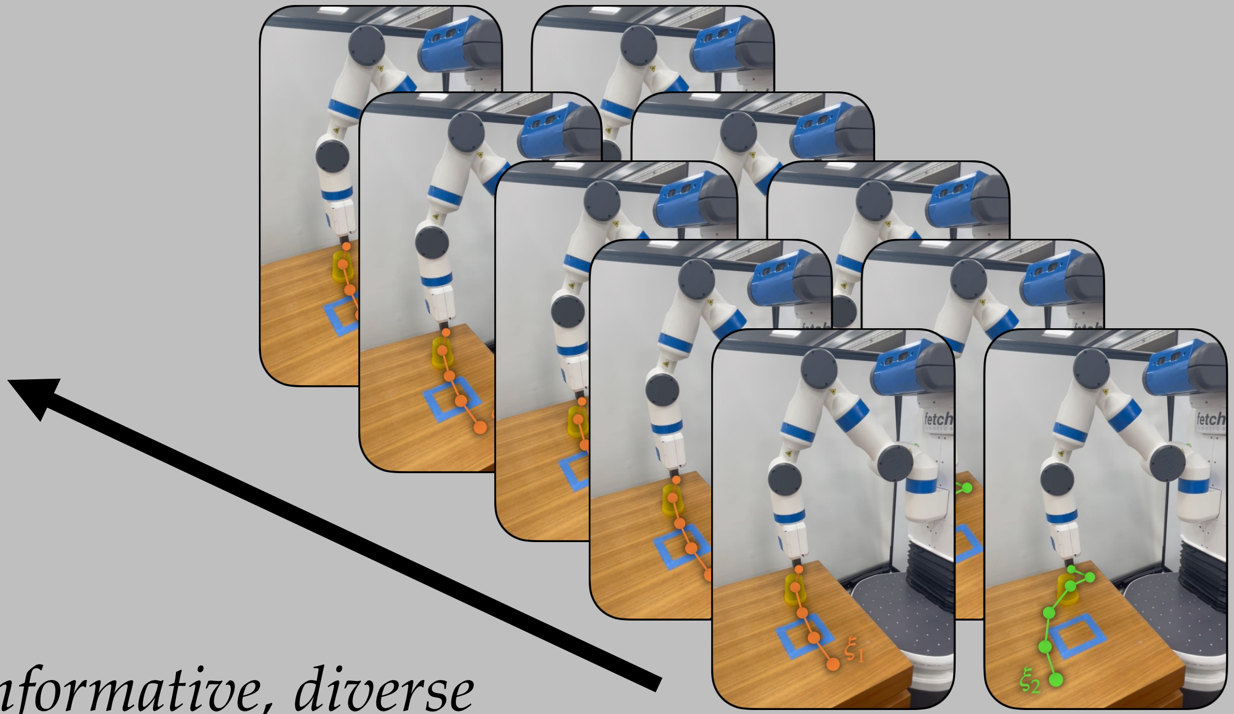
$$R(\xi) = w \cdot \phi(\xi)$$



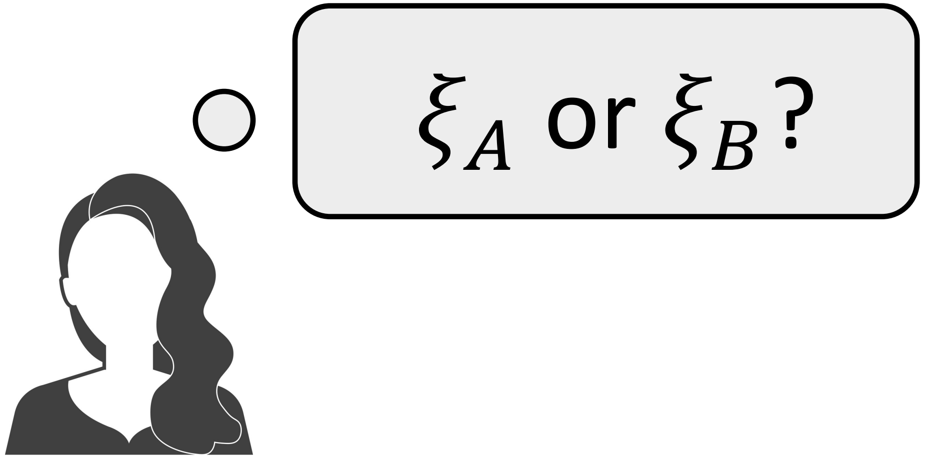
$$R(\xi) = w \cdot \phi(\xi)$$



$$R(\xi) = w \cdot \phi(\xi)$$

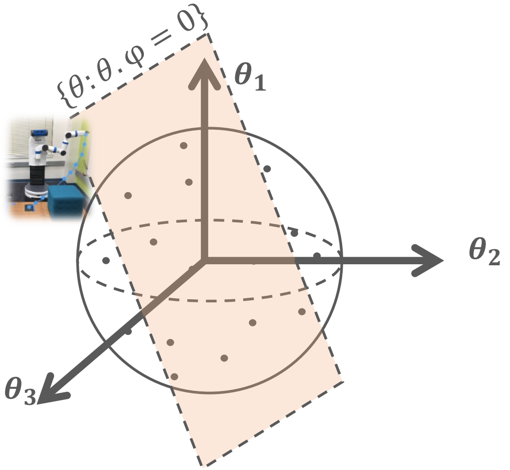


Most informative, diverse sequence of queries



Actively synthesizing queries


Erdem Biyik



minimum volume removed

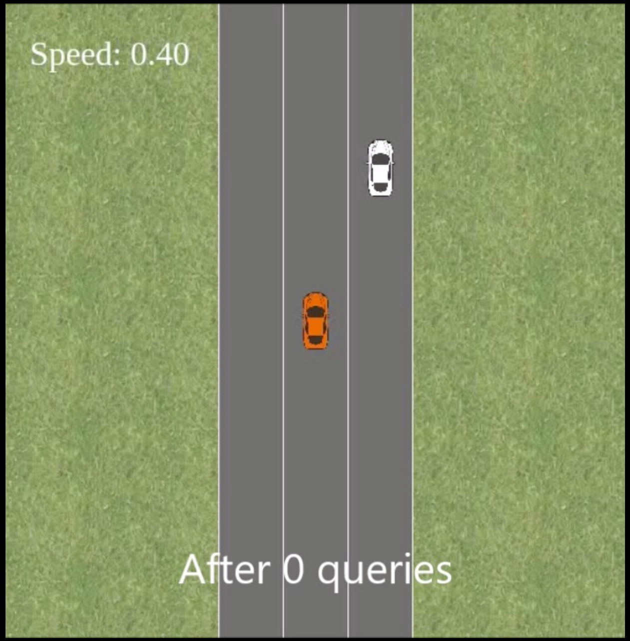
$$\max_{\varphi} \underbrace{\min\{\mathbb{E}[1 - f_{\varphi}(w)], \mathbb{E}[1 - f_{-\varphi}(w)]\}}_{\text{minimum volume removed}}$$

Subject to $\varphi \in \mathbb{F}$

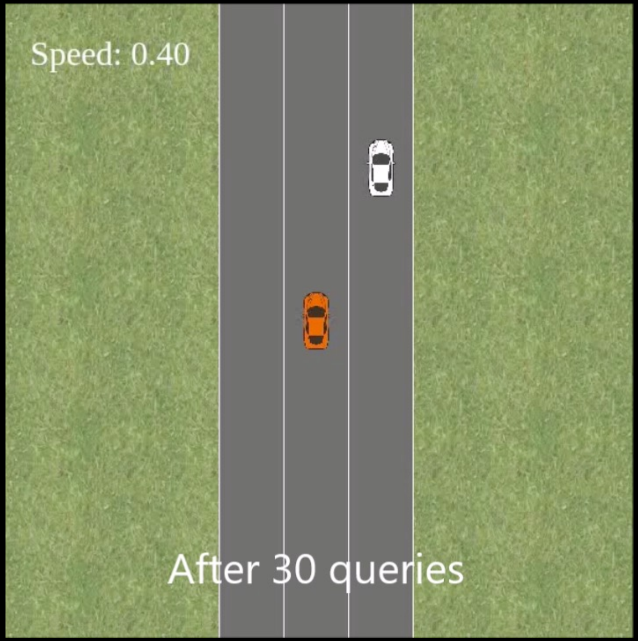
$$\mathbb{F} = \{\varphi : \varphi = \Phi(\xi_A) - \Phi(\xi_B), \xi_A, \xi_B \in \Xi\}$$

Human update function $\quad f_{\varphi}(\boldsymbol{w}) = \min(1, \exp(I_t \boldsymbol{w}^{\top} \varphi))$

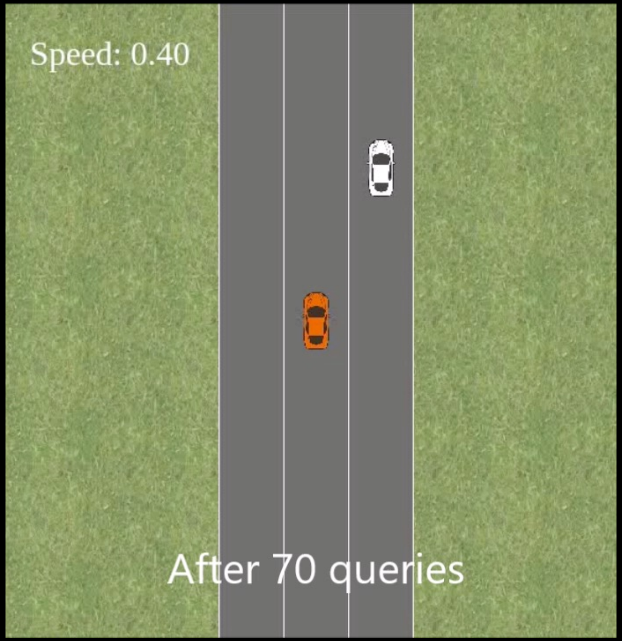
[Sadigh et al. RSS17]
[Biyik et al. CoRL18]
[Biyik et al. CDC19]
[Palan et al. RSS19]
[Biyik et al. CoRL19]
[Basu et al. IROS19]
[Biyik et al. RSS20]
[Myers et al. CoRL21]
[Myers et al. ICRA22]



No prior preference



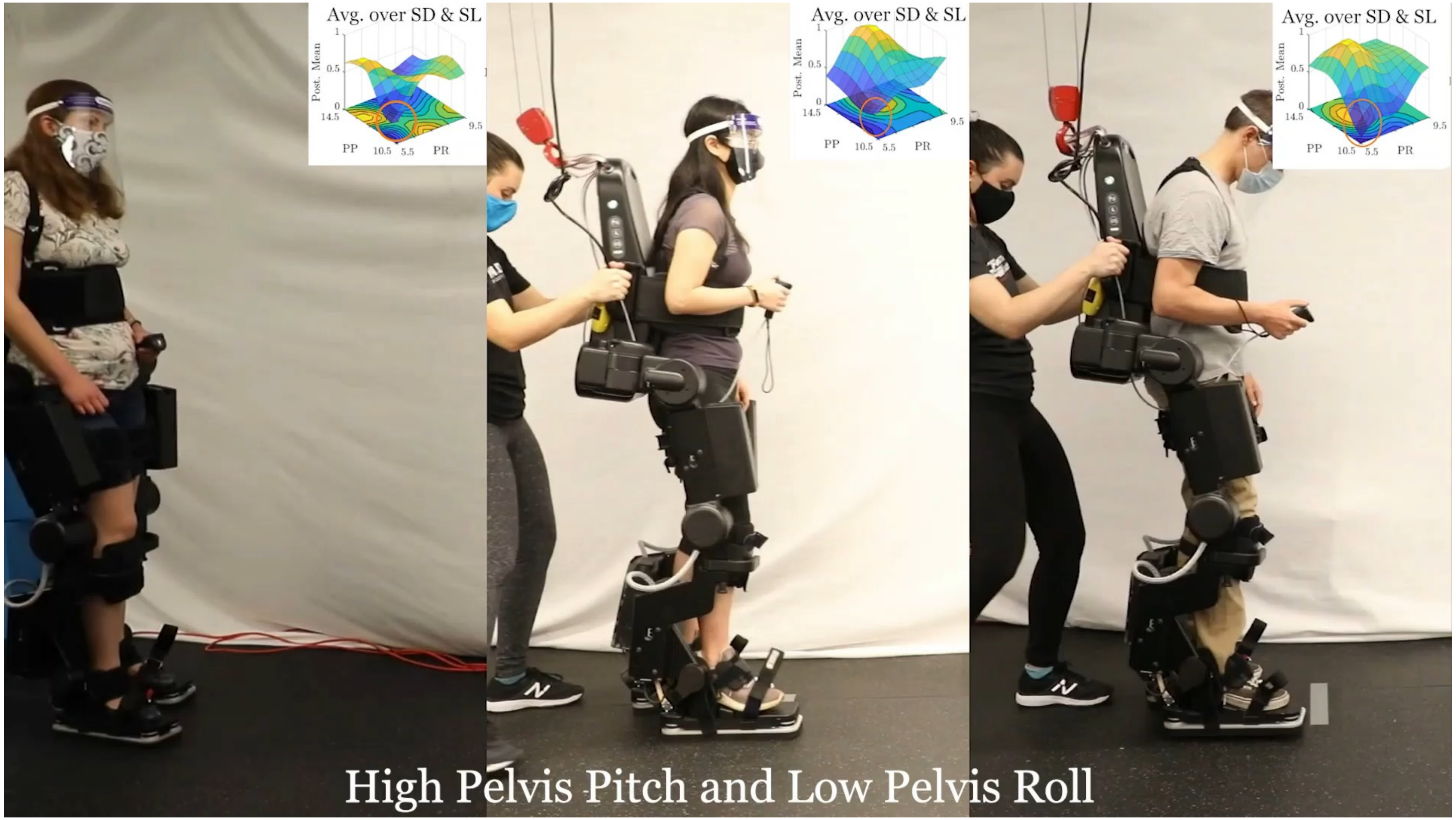
Learns *heading* preferences



Learns *collision avoidance* preferences

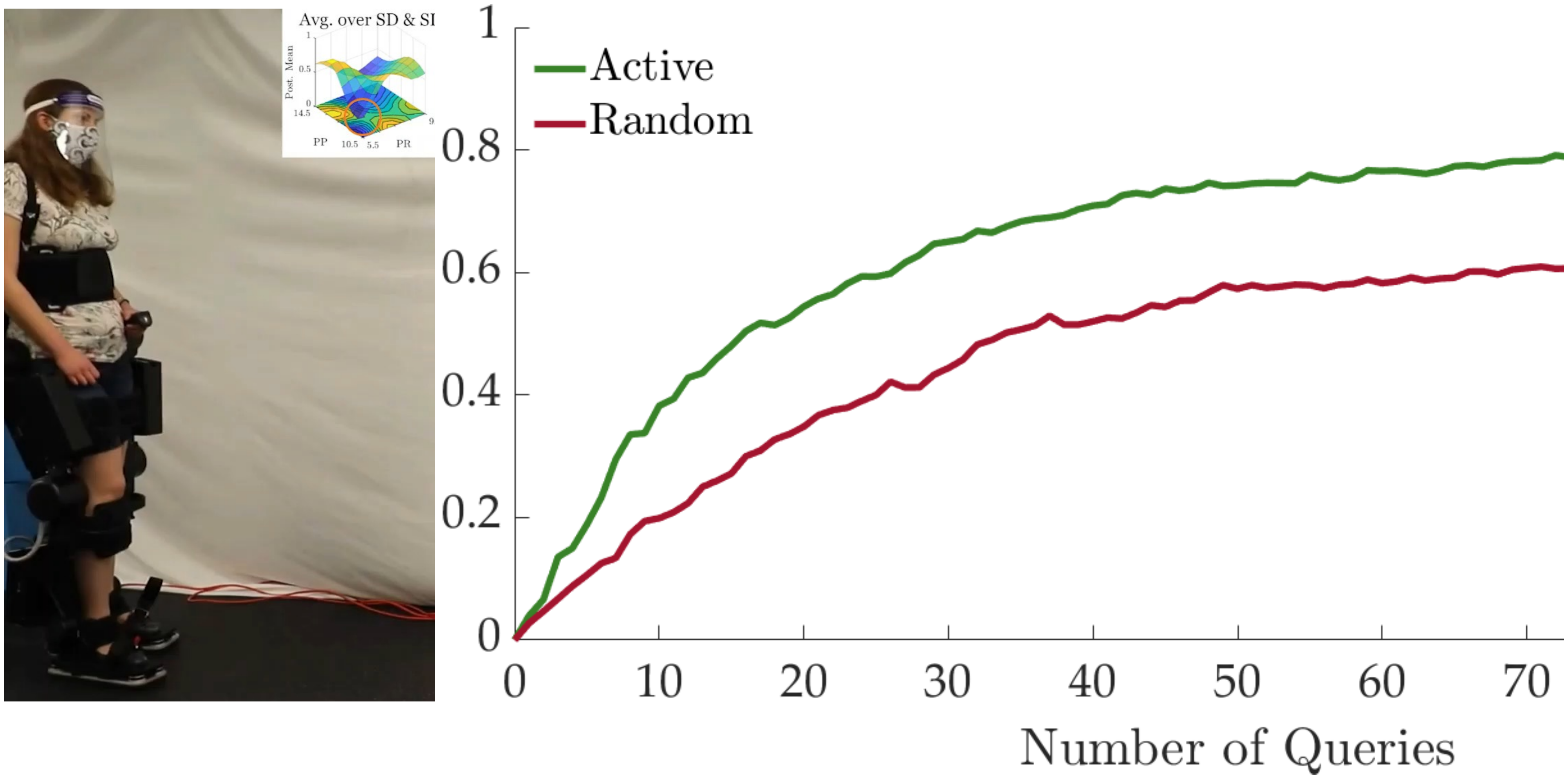
[*Biyik, Sadigh*. CoRL18]

Nonlinear Rewards for Exoskeletons

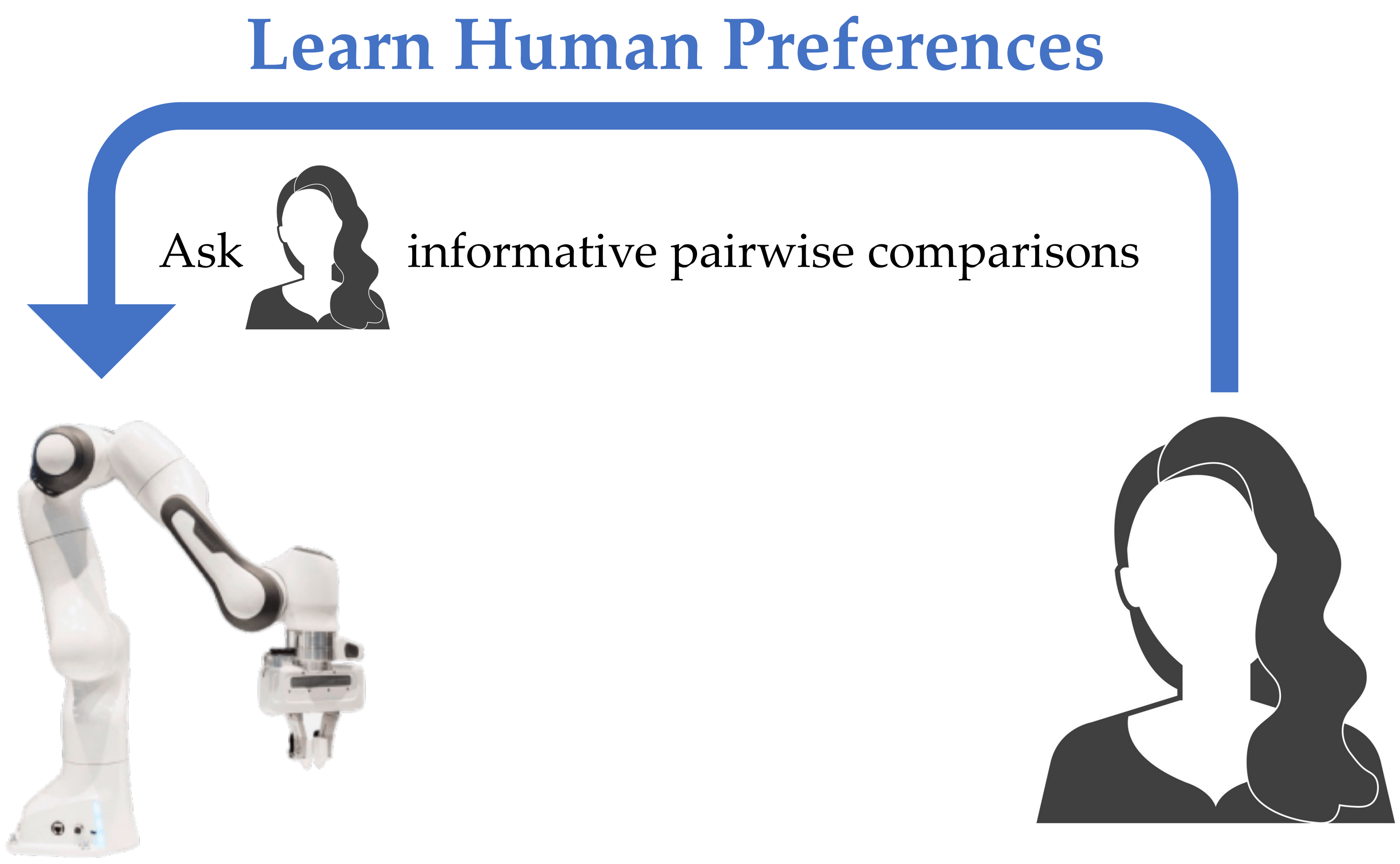


ROIAL: Region of Interest Active Learning for Characterizing Exoskeleton Gait Preference Landscapes
K. Li, et al. ICRA'21.

Nonlinear Rewards for Exoskeletons



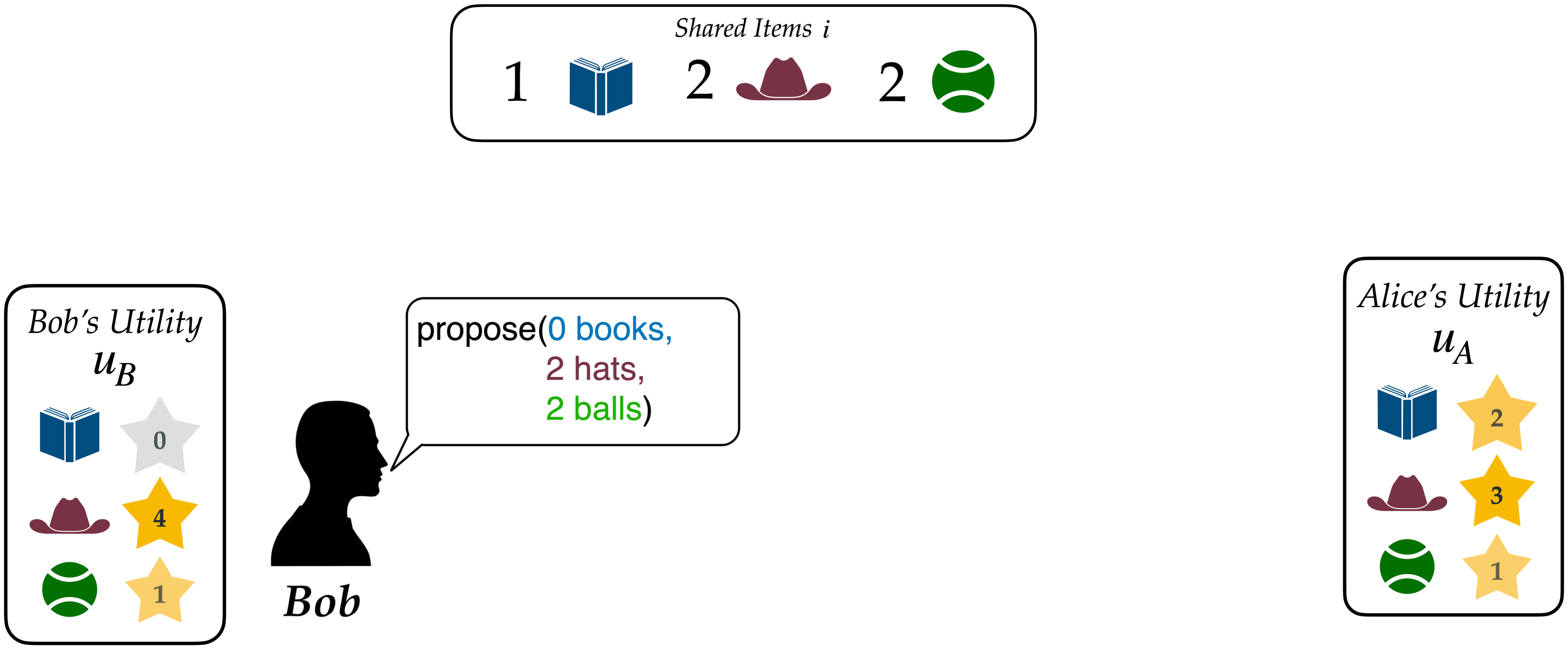
ROIAL: Region of Interest Active Learning for Characterizing Exoskeleton Gait Preference Landscapes
K. Li, et al. ICRA'21.


Learn Human Preferences
Ask
informative pairwise comparisons

Negotiation Domain

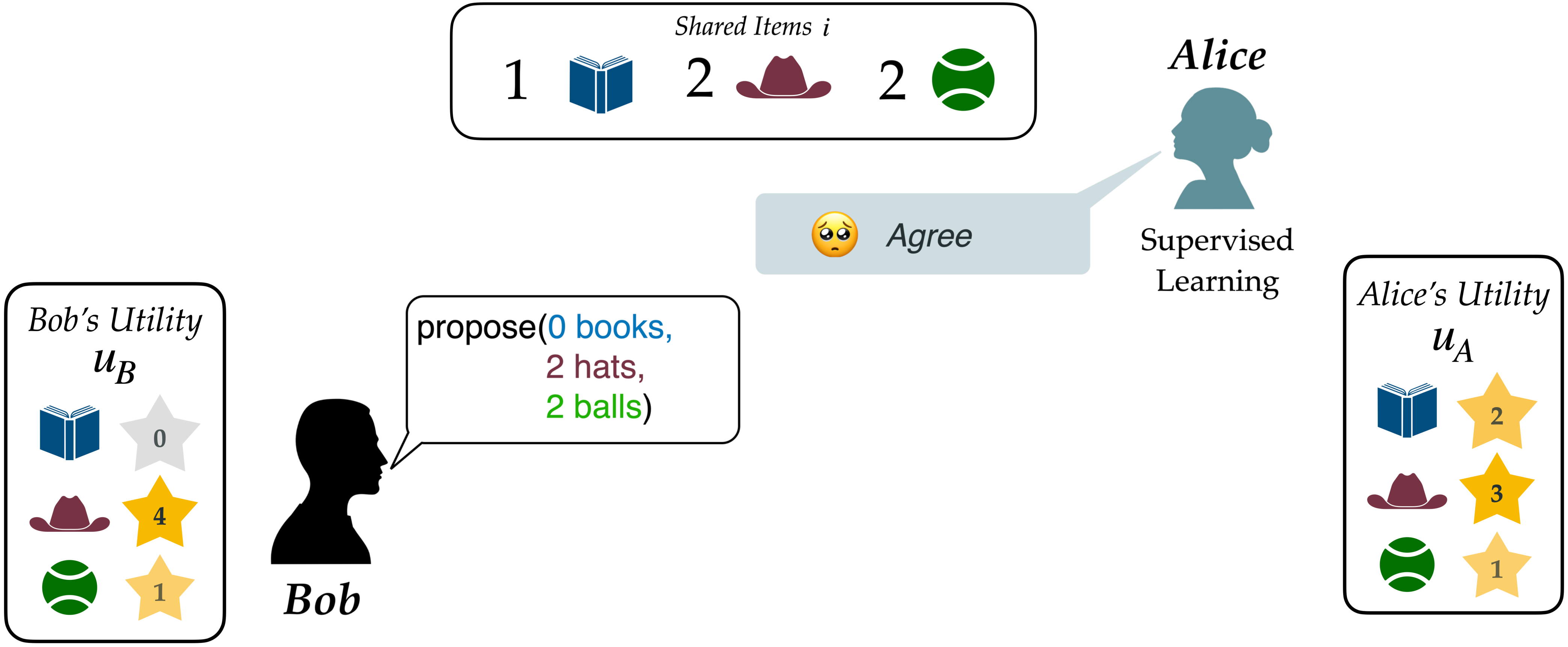
Negotiation Domain

Negotiation Domain

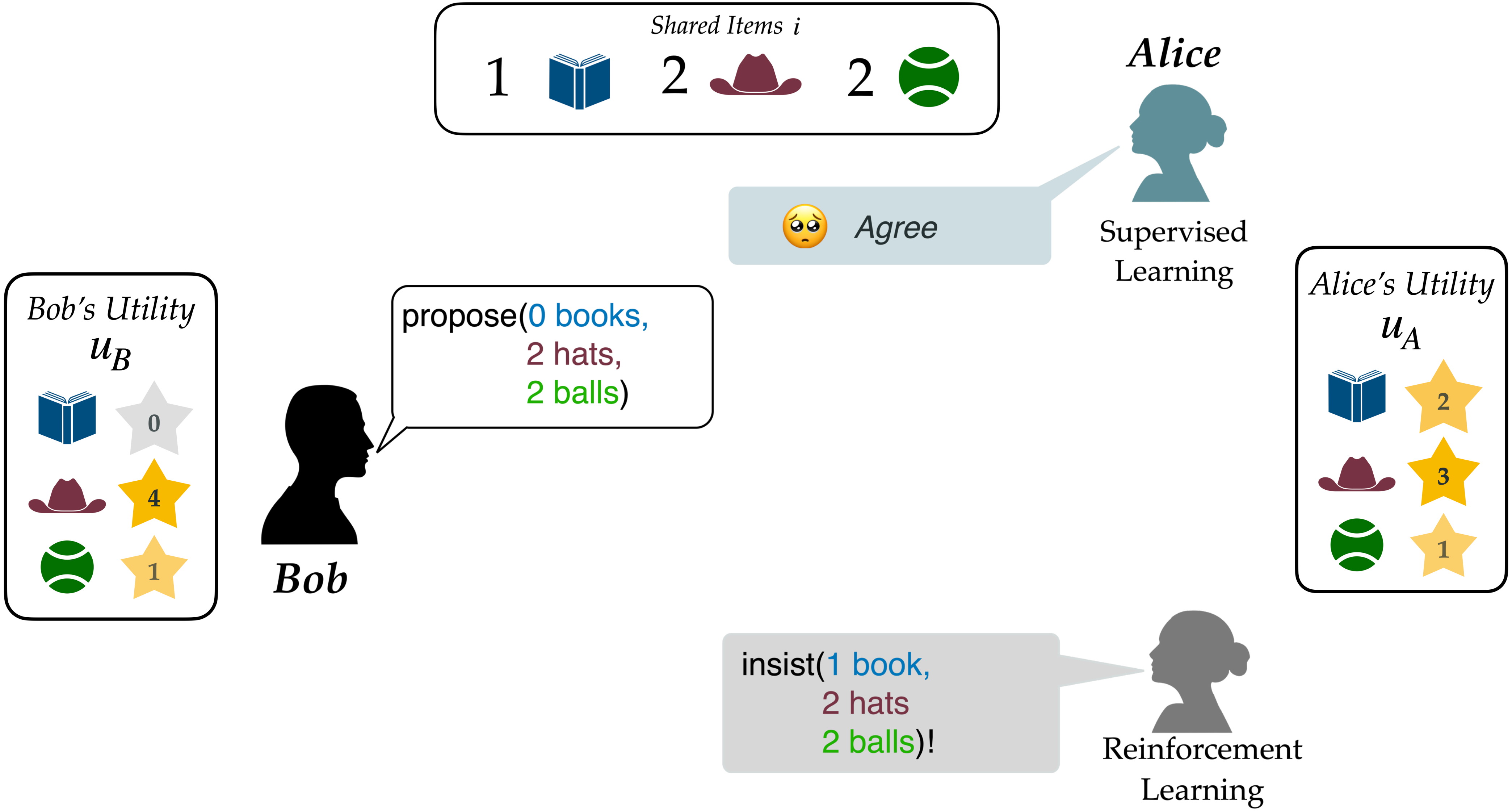


Lewis, Mike, et al. "Deal or no deal? end-to-end learning for negotiation dialogues."

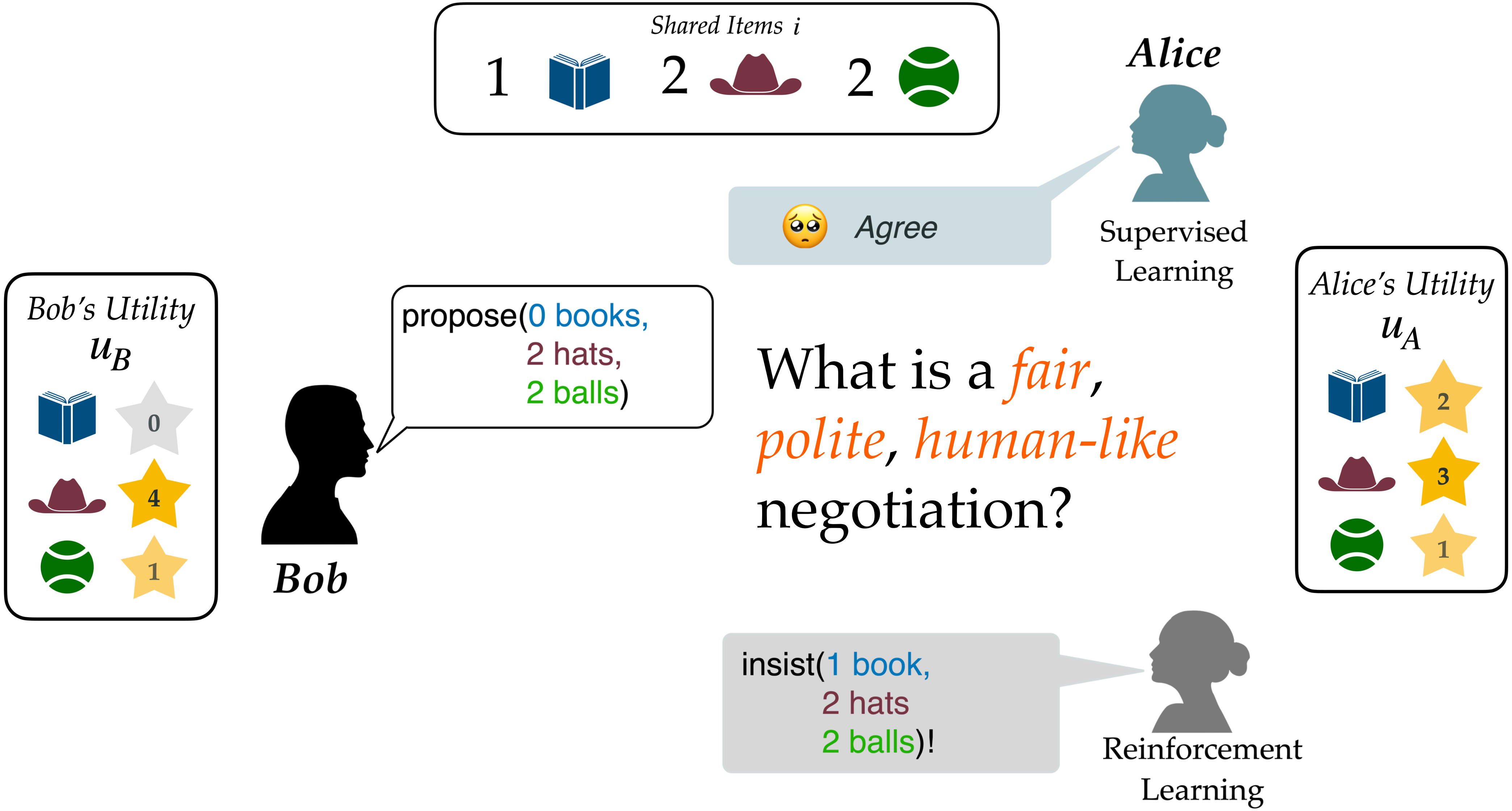
Negotiation Domain



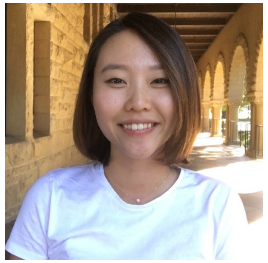
Negotiation Domain



Negotiation Domain

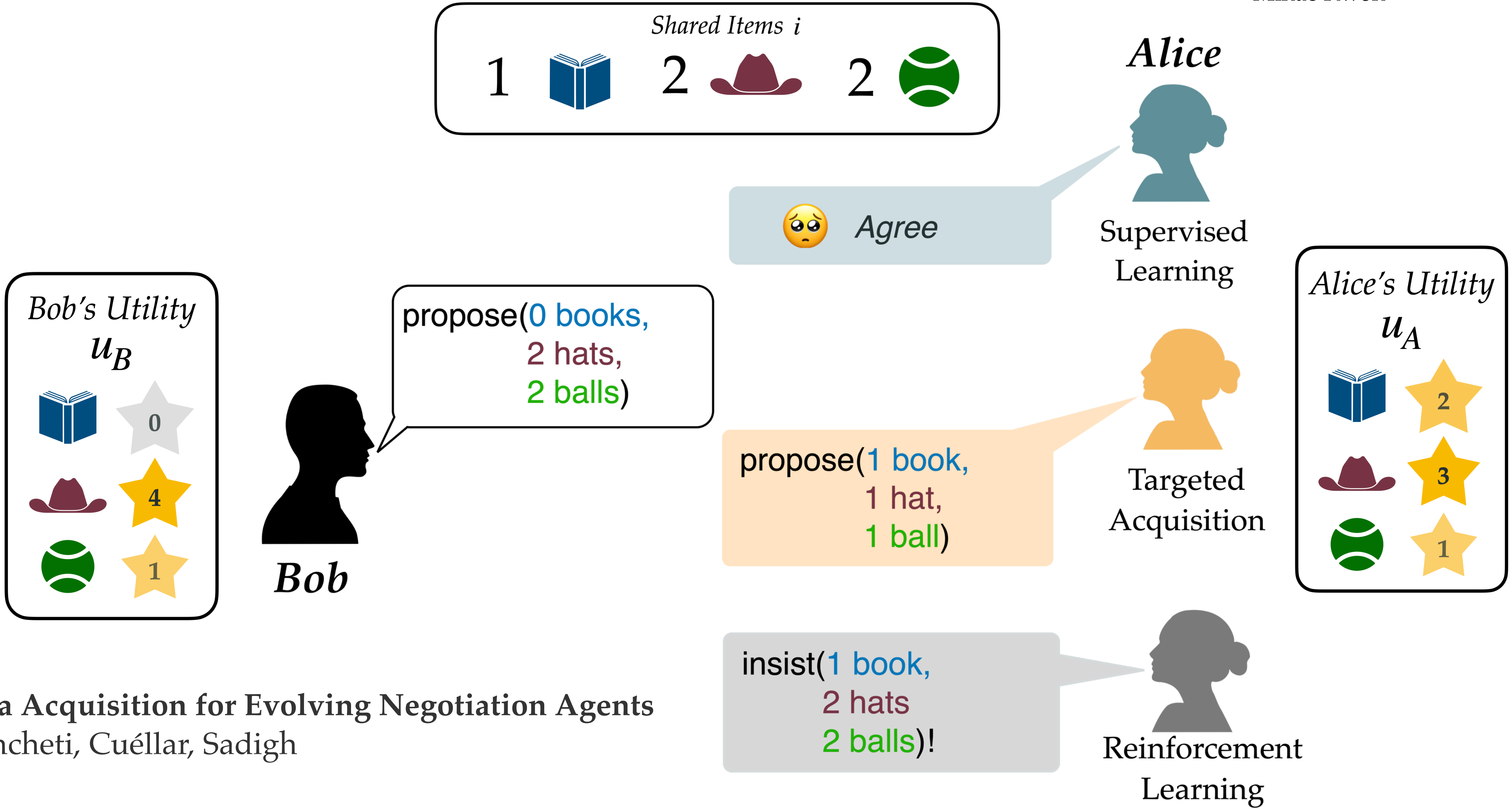


Negotiation Domain

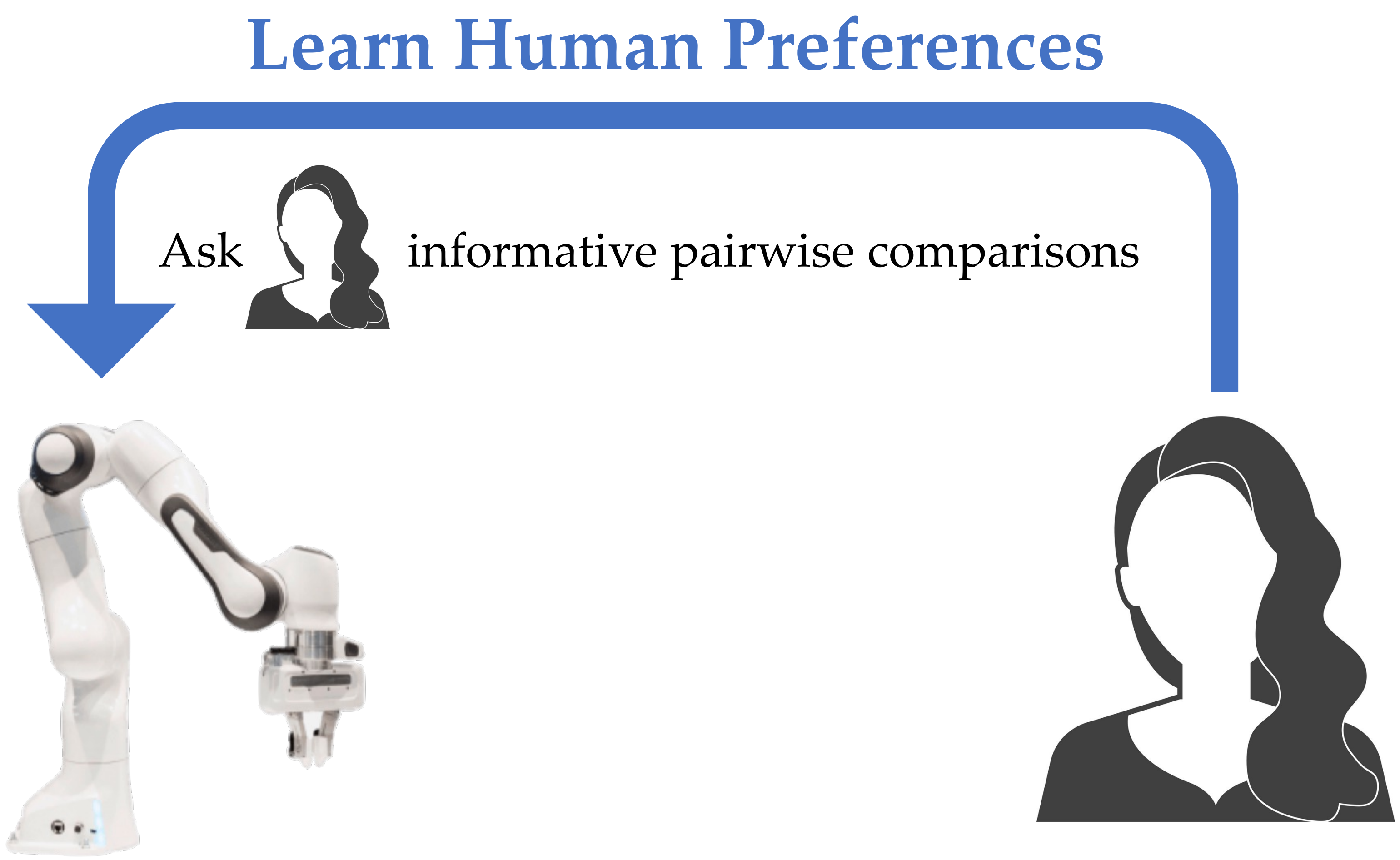


Minae Kwon

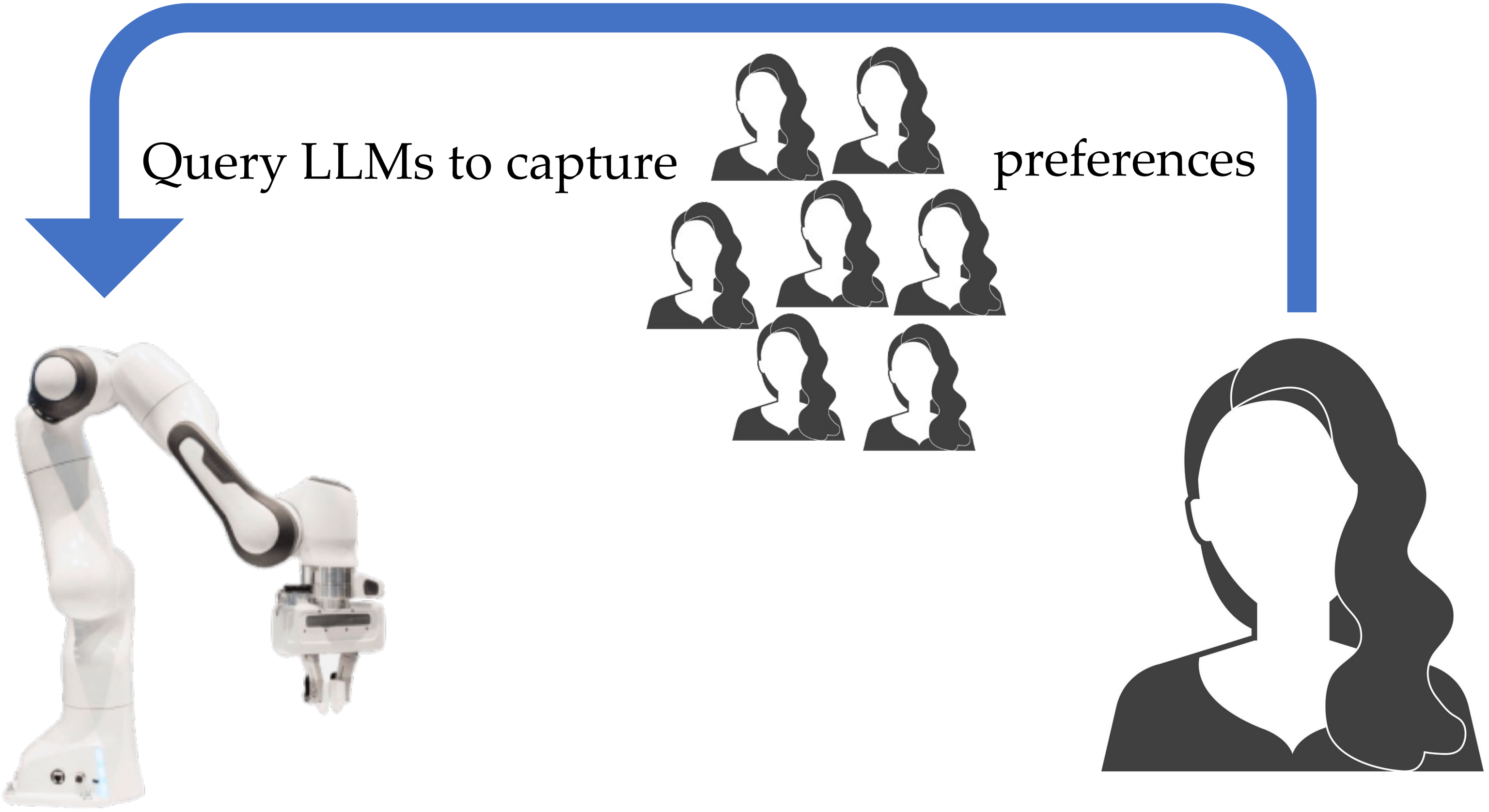
Sidd Karamcheti



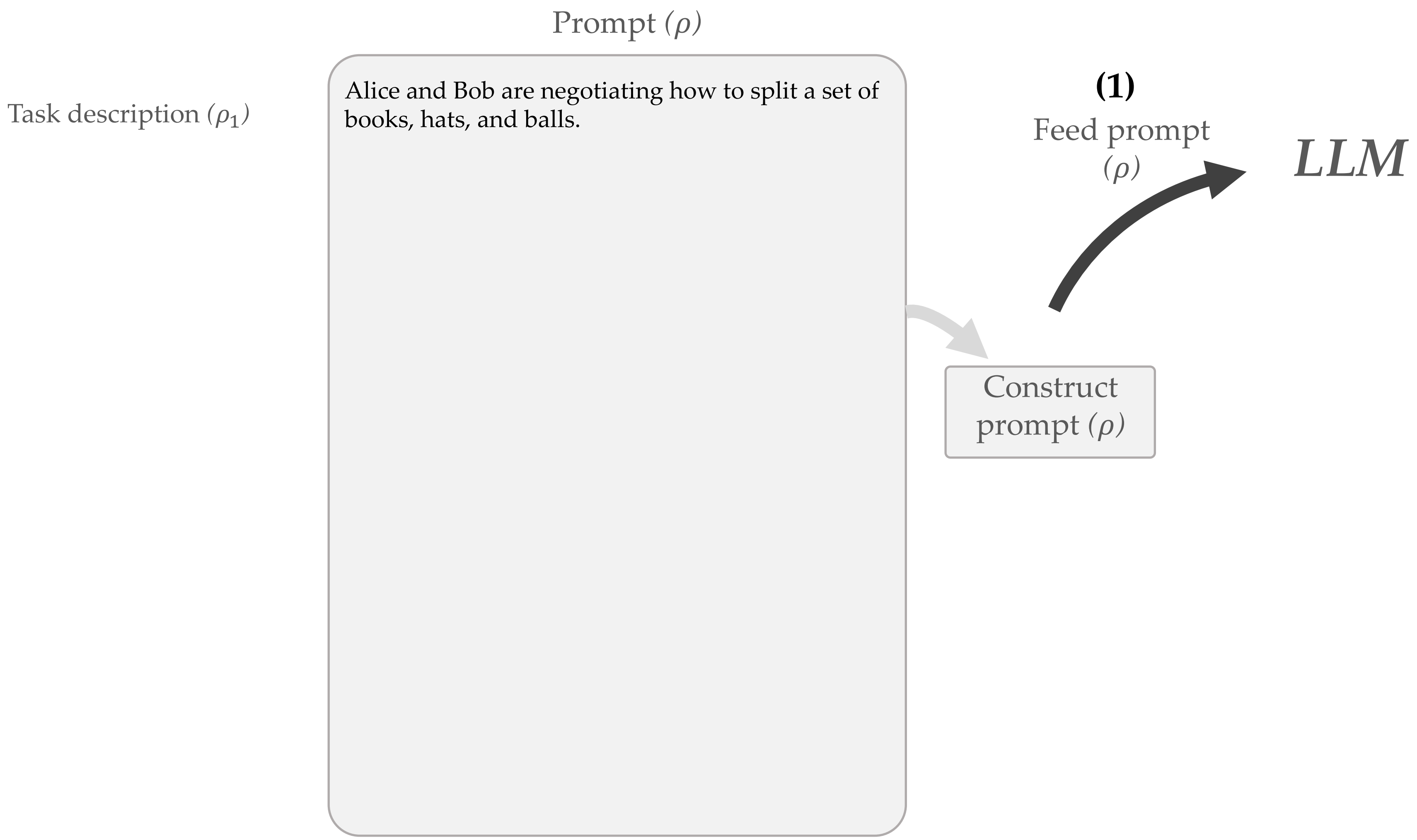
Targeted Data Acquisition for Evolving Negotiation Agents
Kwon, Karamcheti, Cuéllar, Sadigh
ICML 2021


Learn Human Preferences
Ask informative pairwise comparisons

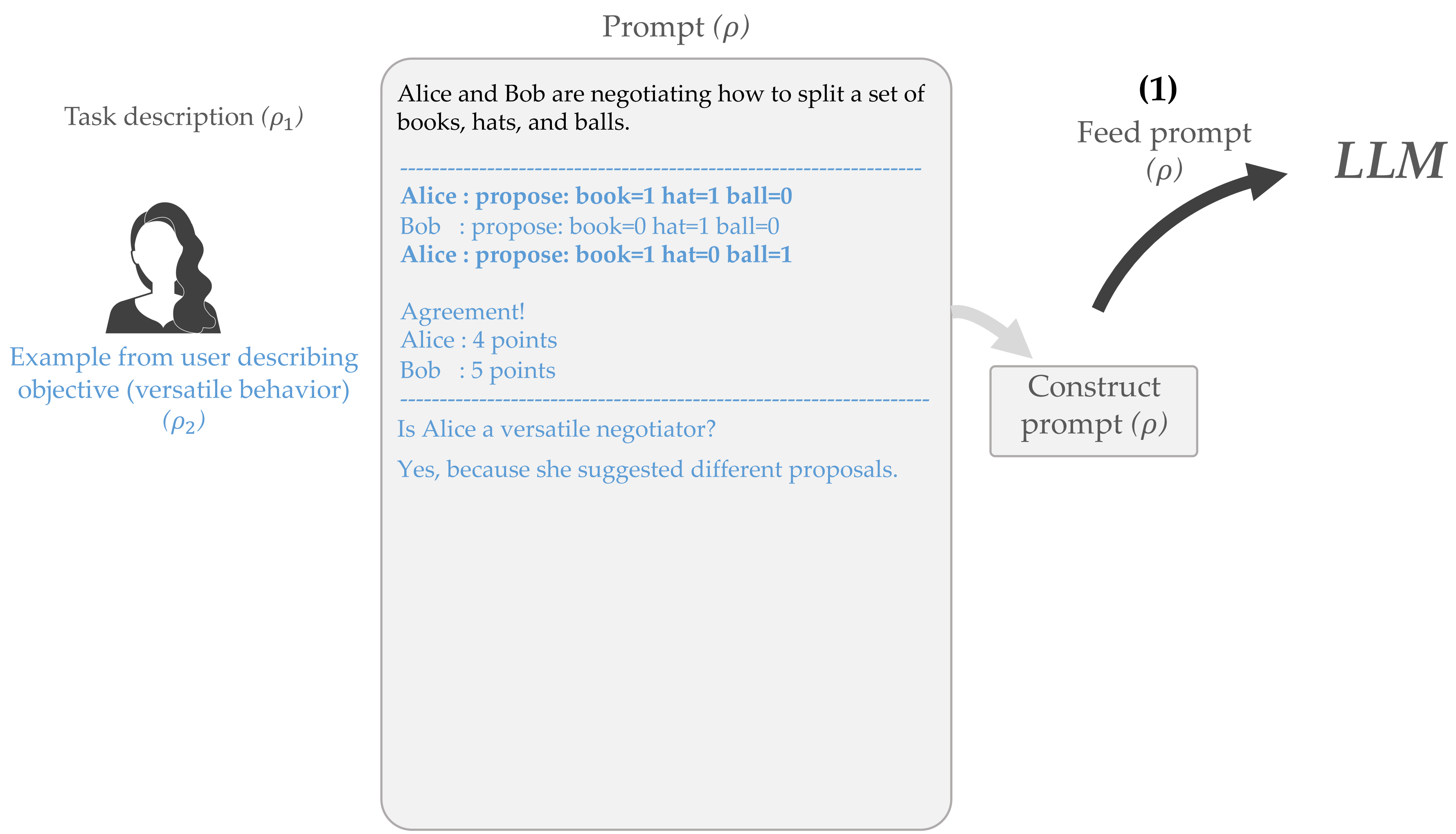
Learn Human Preferences

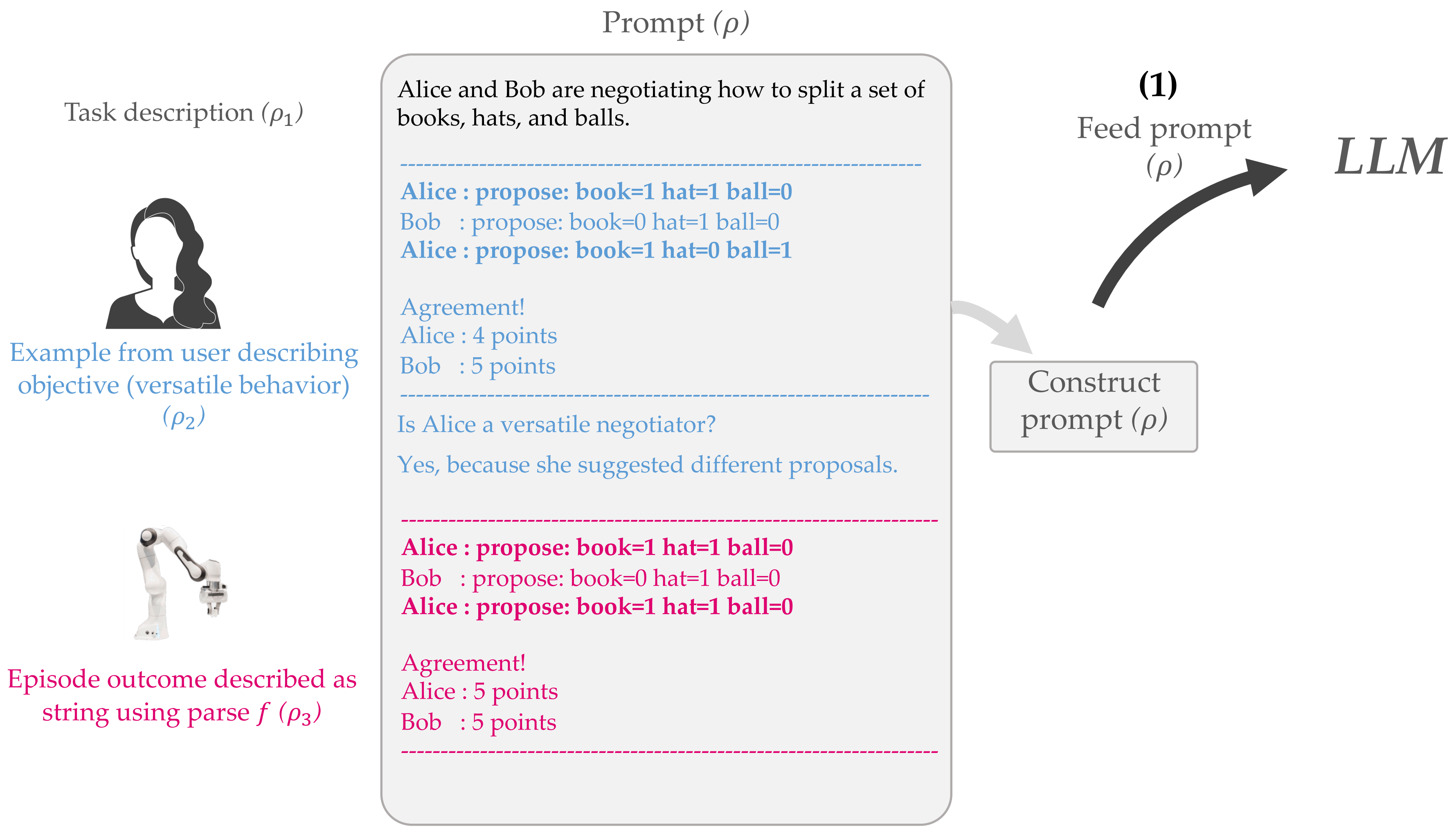


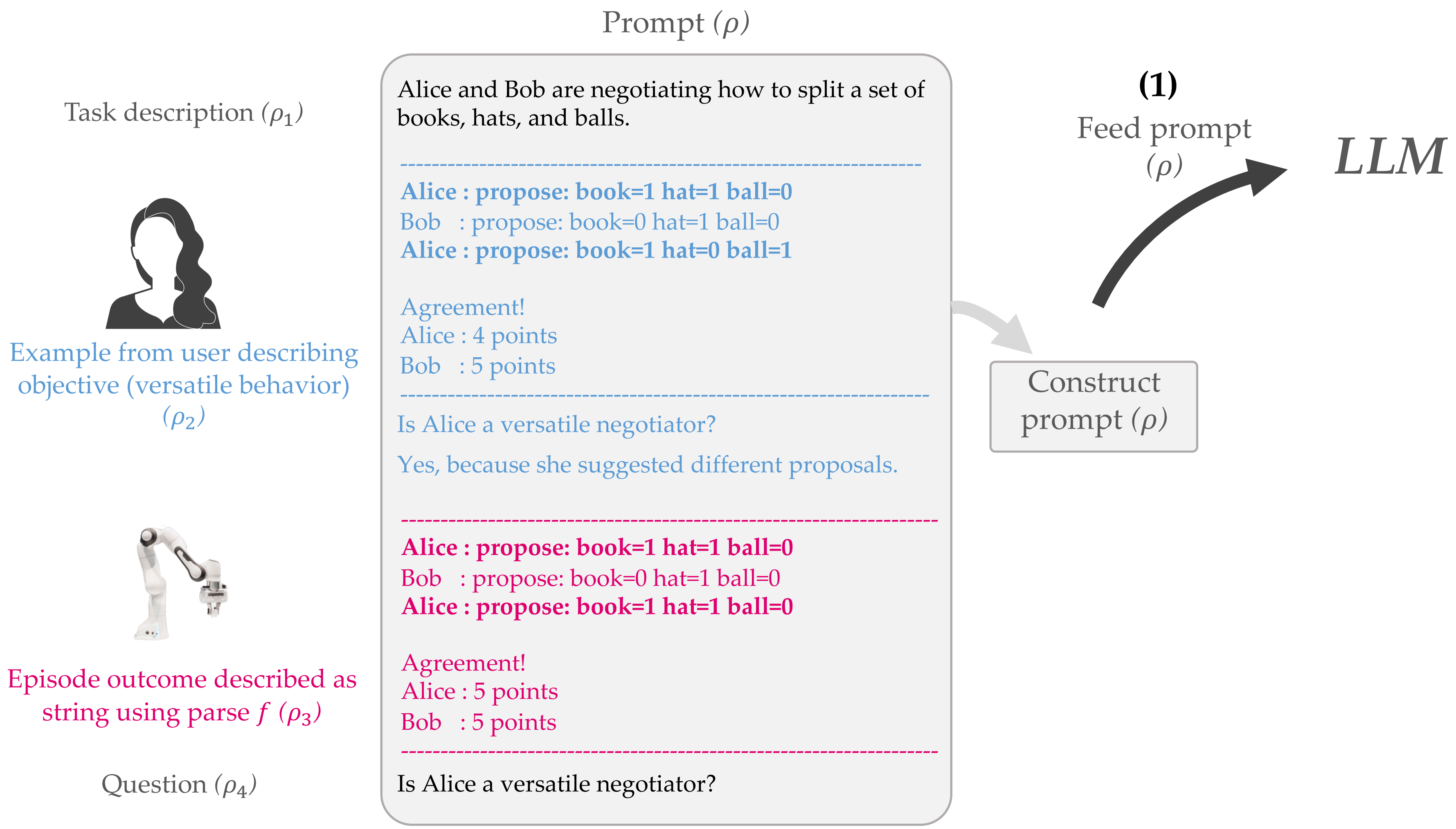
We use LLMs as a proxy reward function
to train RL agents from user inputs

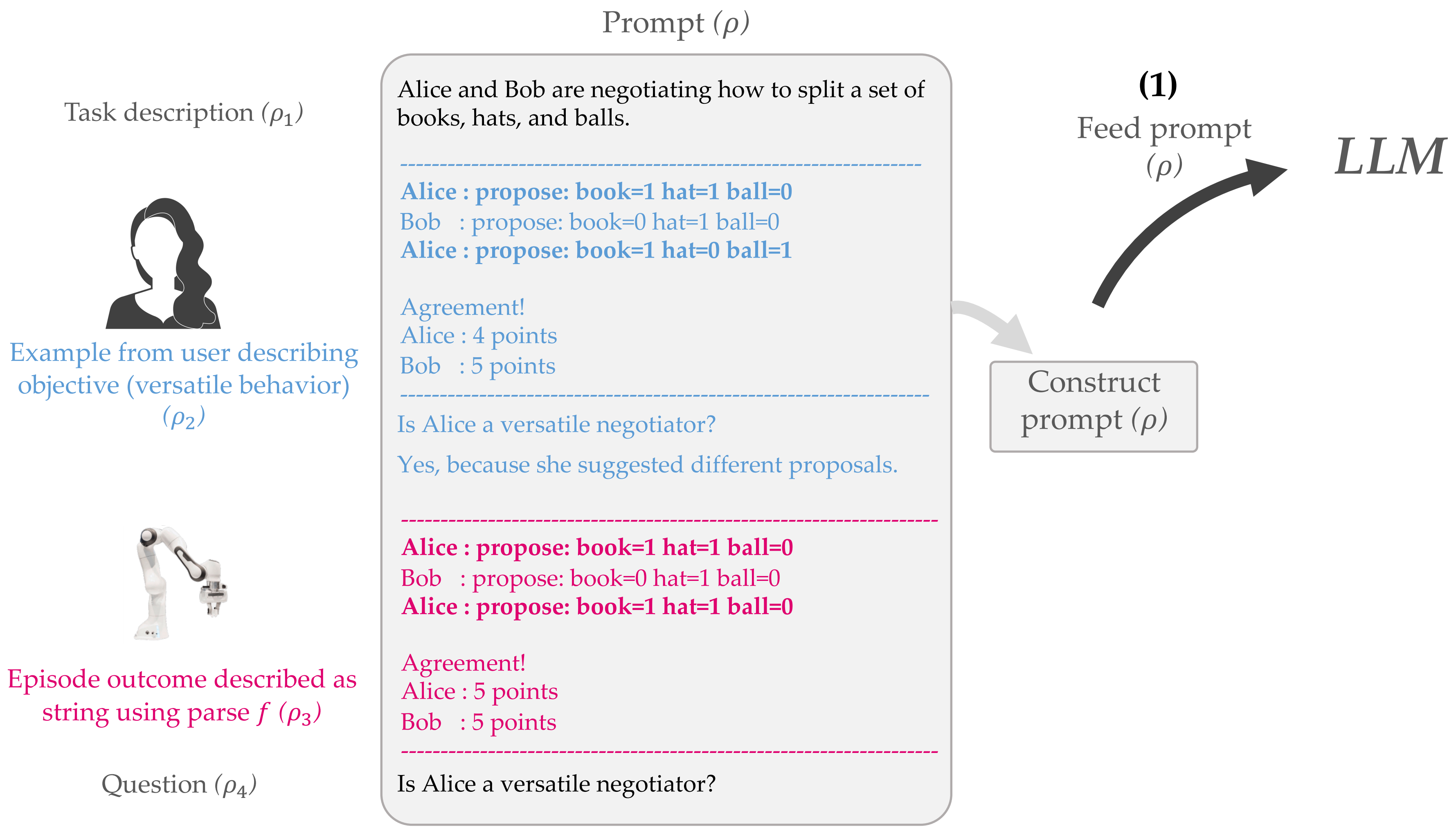


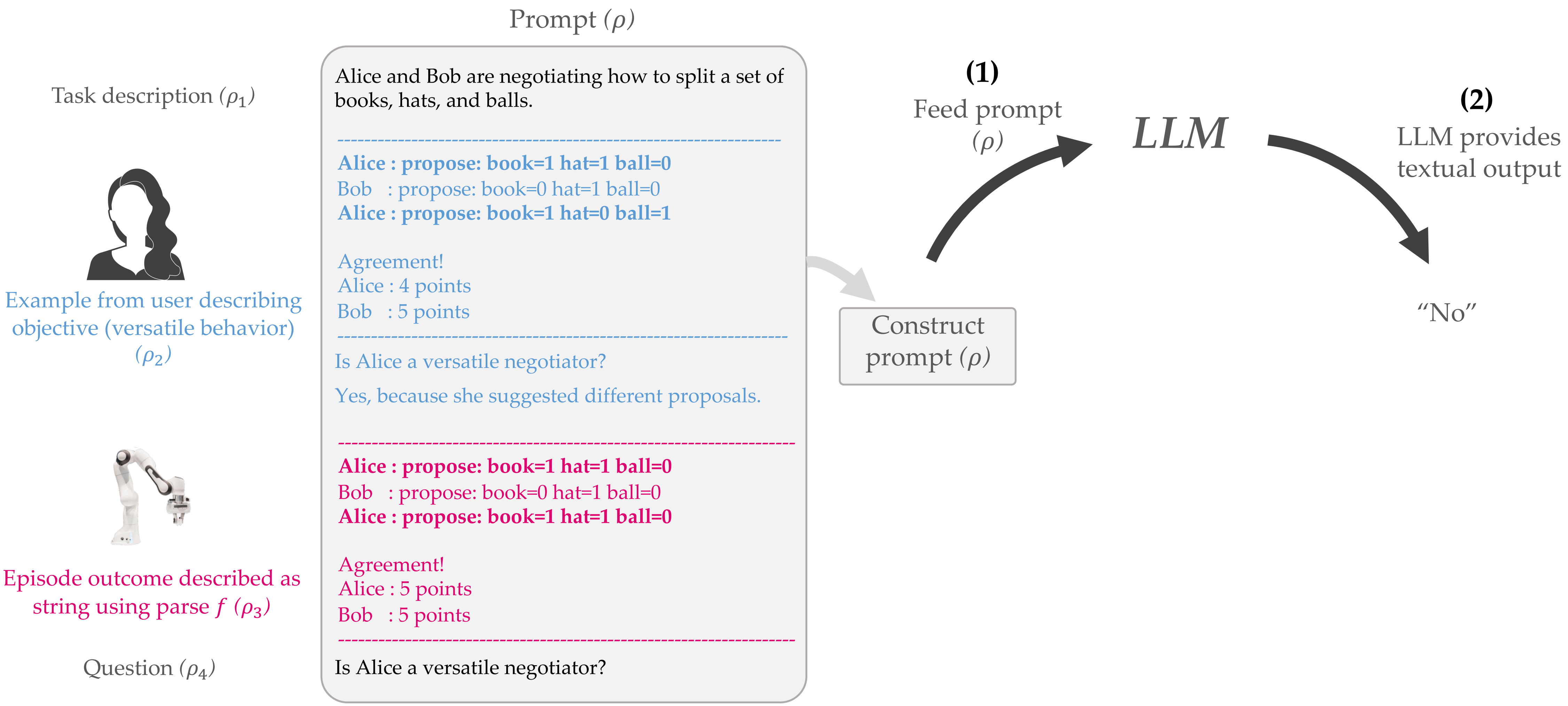
Prompt (ρ)
Task description (ρ_1)
Alice and Bob are negotiating how to split a set of books, hats, and balls.
(1)
Feed prompt (ρ)
LLM
Construct prompt (ρ)


Prompt (ρ)
Task description (ρ_1)
Alice and Bob are negotiating how to split a set of books, hats, and balls.
Example from user describing objective (versatile behavior) (ρ_2)
Alice : propose: book=1 hat=1 ball=0
Bob : propose: book=0 hat=1 ball=0
Alice : propose: book=1 hat=0 ball=1
Agreement!
Alice : 4 points
Bob : 5 points
Is Alice a versatile negotiator?
Yes, because she suggested different proposals.
Construct prompt (ρ)
(1)
Feed prompt (ρ)
LLM

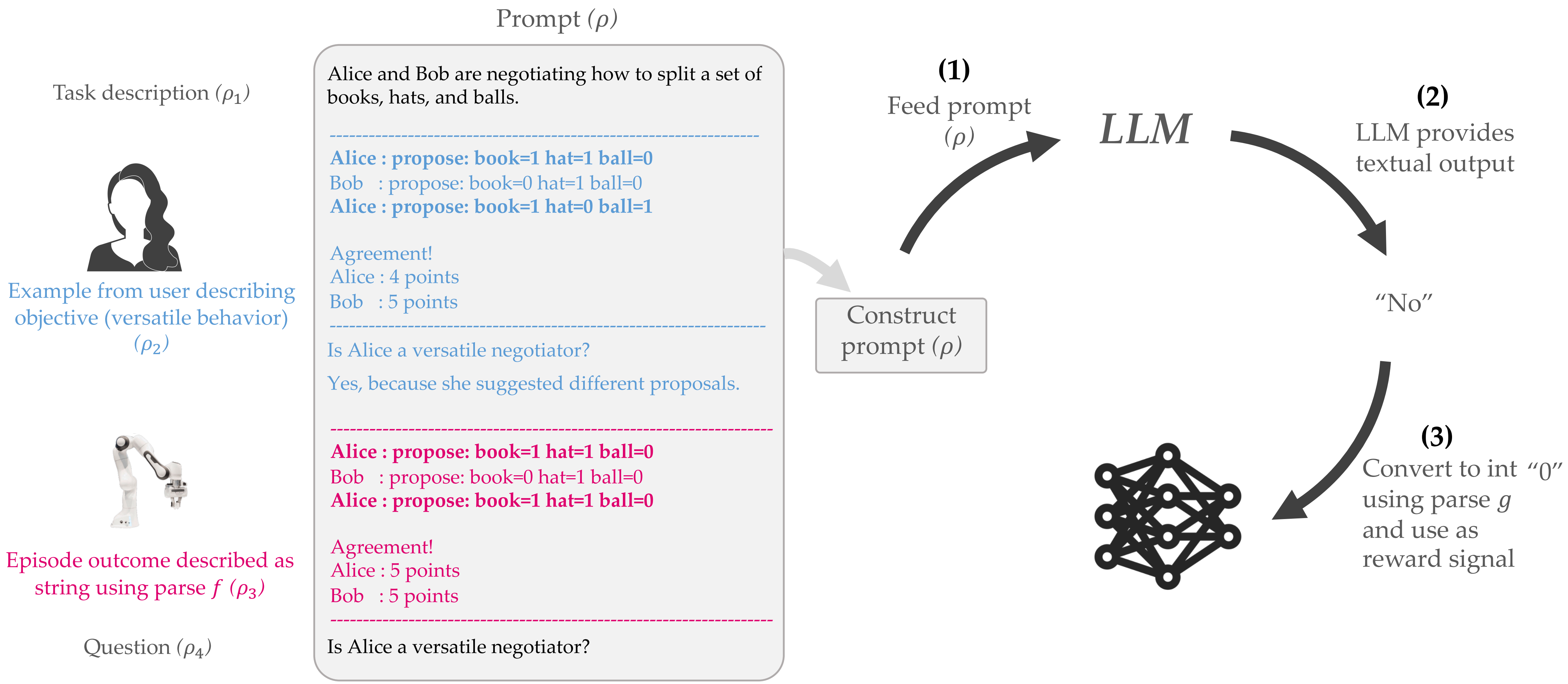

Prompt (ρ)
Task description (ρ_1)
Alice and Bob are negotiating how to split a set of books, hats, and balls.
Alice : propose: book=1 hat=1 ball=0
Bob : propose: book=0 hat=1 ball=0
Alice : propose: book=1 hat=0 ball=1
Agreement!
Alice : 4 points
Bob : 5 points
Is Alice a versatile negotiator?
Yes, because she suggested different proposals.
Example from user describing objective (versatile behavior) (ρ_2)
Alice : propose: book=1 hat=1 ball=0
Bob : propose: book=0 hat=1 ball=0
Alice : propose: book=1 hat=1 ball=0
Agreement!
Alice : 5 points
Bob : 5 points
Episode outcome described as string using parse f (ρ_3)
Construct prompt (ρ)
(1)
Feed prompt (ρ)
LLM

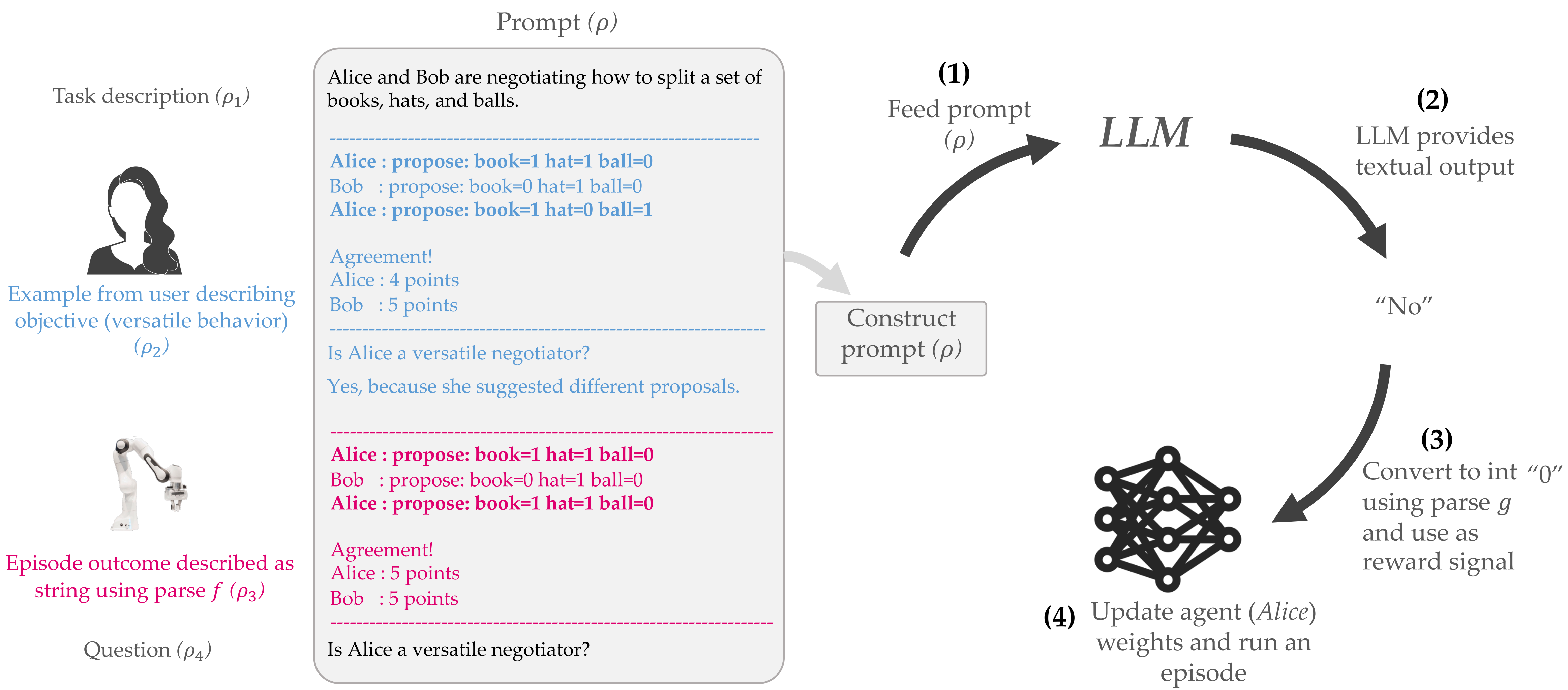

Prompt (ρ)
Task description (ρ_1)
Alice and Bob are negotiating how to split a set of books, hats, and balls.
Example from user describing objective (versatile behavior) (ρ_2)
Alice : propose: book=1 hat=1 ball=0
Bob : propose: book=0 hat=1 ball=0
Alice : propose: book=1 hat=0 ball=1
Agreement!
Alice : 4 points
Bob : 5 points
Is Alice a versatile negotiator?
Yes, because she suggested different proposals.
Episode outcome described as string using parse f (ρ_3)
Alice : propose: book=1 hat=1 ball=0
Bob : propose: book=0 hat=1 ball=0
Alice : propose: book=1 hat=1 ball=0
Agreement!
Alice : 5 points
Bob : 5 points
Question (ρ_4)
Is Alice a versatile negotiator?
Construct prompt (ρ)
(1)
Feed prompt (ρ)
LLM

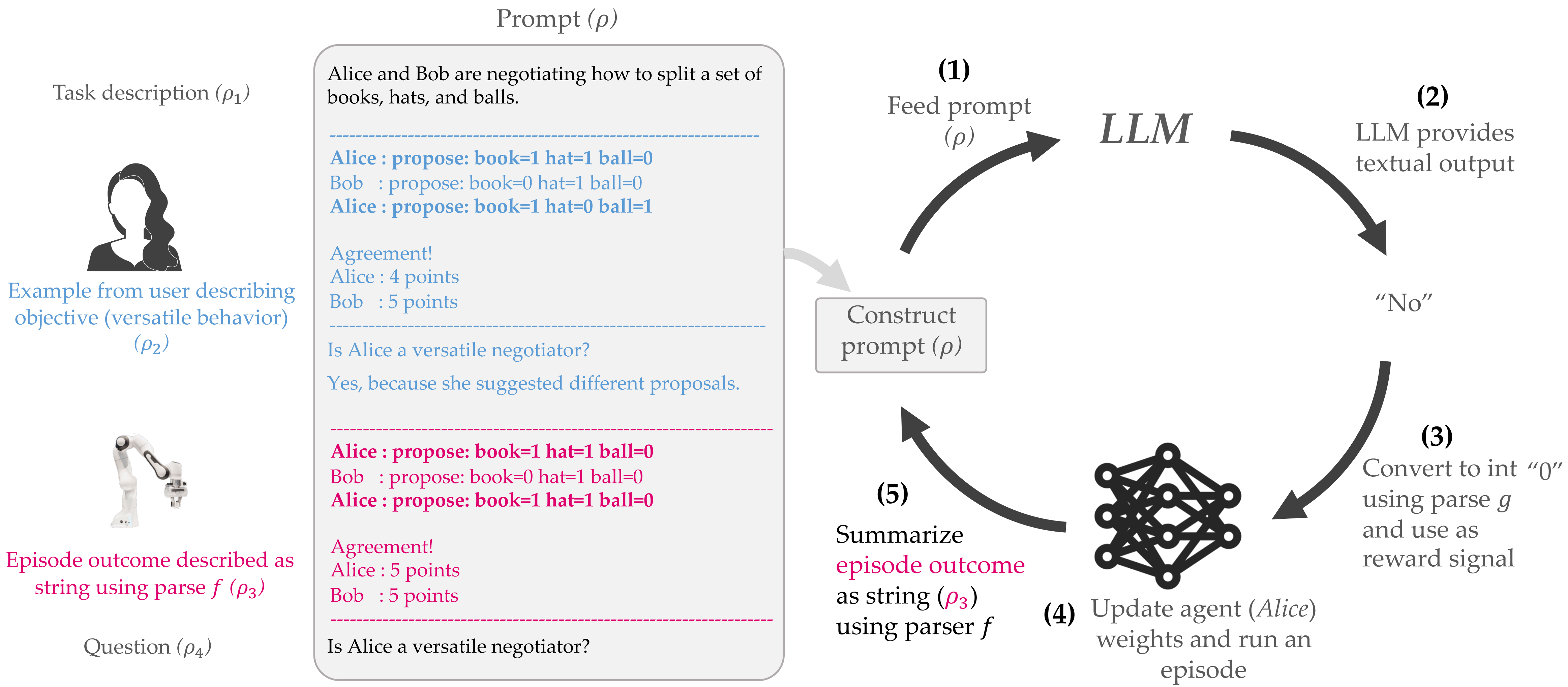

Prompt (ρ)
Task description (ρ_1)
Alice and Bob are negotiating how to split a set of books, hats, and balls.
Example from user describing objective (versatile behavior) (ρ_2)
Alice : propose: book=1 hat=1 ball=0
Bob : propose: book=0 hat=1 ball=0
Alice : propose: book=1 hat=0 ball=1
Agreement!
Alice : 4 points
Bob : 5 points
Is Alice a versatile negotiator?
Yes, because she suggested different proposals.
Episode outcome described as string using parse f (ρ_3)
Alice : propose: book=1 hat=1 ball=0
Bob : propose: book=0 hat=1 ball=0
Alice : propose: book=1 hat=1 ball=0
Agreement!
Alice : 5 points
Bob : 5 points
Question (ρ_4)
Is Alice a versatile negotiator?
Construct prompt (ρ)
(1)
Feed prompt (ρ)
LLM



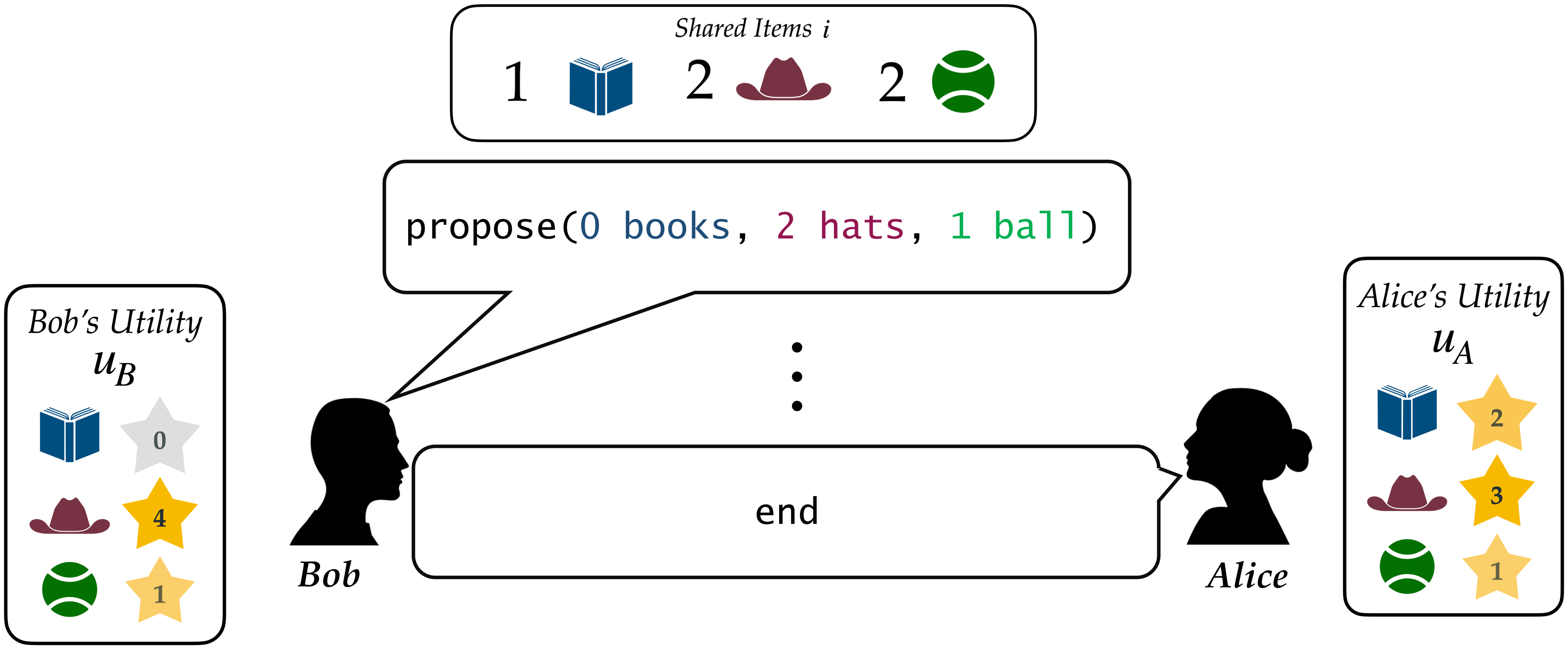
Prompt (ρ)
Task description (ρ_1)
Alice and Bob are negotiating how to split a set of books, hats, and balls.
Example from user describing objective (versatile behavior) (ρ_2)
Alice : propose: book=1 hat=1 ball=0
Bob : propose: book=0 hat=1 ball=0
Alice : propose: book=1 hat=0 ball=1
Agreement!
Alice : 4 points
Bob : 5 points
Is Alice a versatile negotiator?
Yes, because she suggested different proposals.
Episode outcome described as string using parse f (ρ_3)
Alice : propose: book=1 hat=1 ball=0
Bob : propose: book=0 hat=1 ball=0
Alice : propose: book=1 hat=1 ball=0
Agreement!
Alice : 5 points
Bob : 5 points
Question (ρ_4)
Is Alice a versatile negotiator?
Construct prompt (ρ)
(1)
Feed prompt (ρ)
LLM
(2)
LLM provides textual output
"No"


Prompt (ρ)
Task description (ρ_1)
Alice and Bob are negotiating how to split a set of books, hats, and balls.
Alice : propose: book=1 hat=1 ball=0
Bob : propose: book=0 hat=1 ball=0
Alice : propose: book=1 hat=0 ball=1
Agreement!
Alice : 4 points
Bob : 5 points
Is Alice a versatile negotiator?
Yes, because she suggested different proposals.
Example from user describing objective (versatile behavior) (ρ_2)
Alice : propose: book=1 hat=1 ball=0
Bob : propose: book=0 hat=1 ball=0
Alice : propose: book=1 hat=1 ball=0
Agreement!
Alice : 5 points
Bob : 5 points
Episode outcome described as string using parse f (ρ_3)
Question (ρ_4)
Is Alice a versatile negotiator?
Construct prompt (ρ)
(1)
Feed prompt (ρ)
LLM
(2)
LLM provides textual output
"No"
(3)
Convert to int "0" using parse g and use as reward signal


Prompt (ρ)
Task description (ρ1)
Alice and Bob are negotiating how to split a set of books, hats, and balls.
Alice : propose: book=1 hat=1 ball=0
Bob : propose: book=0 hat=1 ball=0
Alice : propose: book=1 hat=0 ball=1
Agreement!
Alice : 4 points
Bob : 5 points
Is Alice a versatile negotiator?
Yes, because she suggested different proposals.
Example from user describing objective (versatile behavior) (ρ2)
Alice : propose: book=1 hat=1 ball=0
Bob : propose: book=0 hat=1 ball=0
Alice : propose: book=1 hat=1 ball=0
Agreement!
Alice : 5 points
Bob : 5 points
Episode outcome described as string using parse f (ρ3)
Question (ρ4)
Is Alice a versatile negotiator?
Construct prompt (ρ)
(1)
Feed prompt (ρ)
LLM
(2)
LLM provides textual output
"No"
(3)
Convert to int "0" using parse g and use as reward signal
(4)
Update agent (Alice) weights and run an episode


Prompt (ρ)
Task description (ρ_1)
Alice and Bob are negotiating how to split a set of books, hats, and balls.
Alice : propose: book=1 hat=1 ball=0
Bob : propose: book=0 hat=1 ball=0
Alice : propose: book=1 hat=0 ball=1
Agreement!
Alice : 4 points
Bob : 5 points
Is Alice a versatile negotiator?
Yes, because she suggested different proposals.
Example from user describing objective (versatile behavior) (ρ_2)
Alice : propose: book=1 hat=1 ball=0
Bob : propose: book=0 hat=1 ball=0
Alice : propose: book=1 hat=1 ball=0
Agreement!
Alice : 5 points
Bob : 5 points
Episode outcome described as string using parse f (ρ_3)
Question (ρ_4)
Is Alice a versatile negotiator?
Construct prompt (ρ)
(1)
Feed prompt (ρ)
LLM
(2)
LLM provides textual output
"No"
(3)
Convert to int "0" using parse g and use as reward signal
(4)
Update agent (Alice) weights and run an episode
(5)
Summarize episode outcome as string (ρ_3) using parser f

DEALORNODEAL Negotiation Task



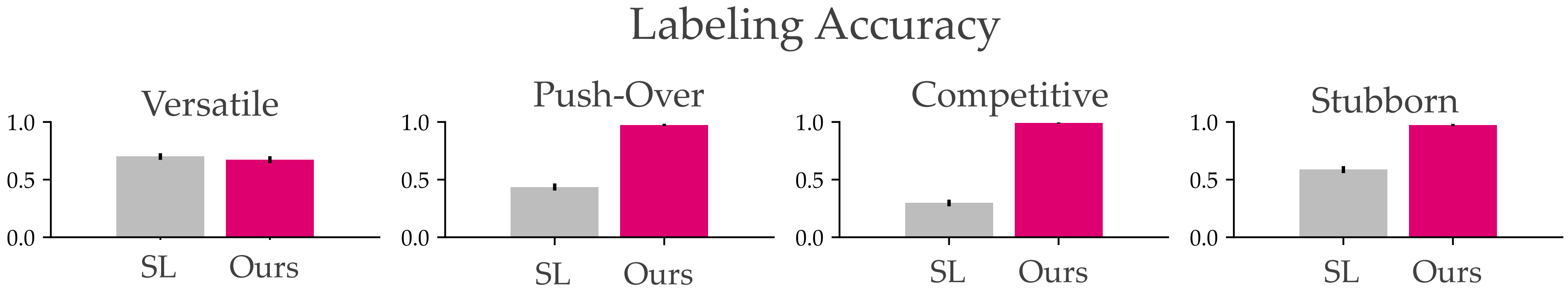
DealOrNoDeal Negotiation Task

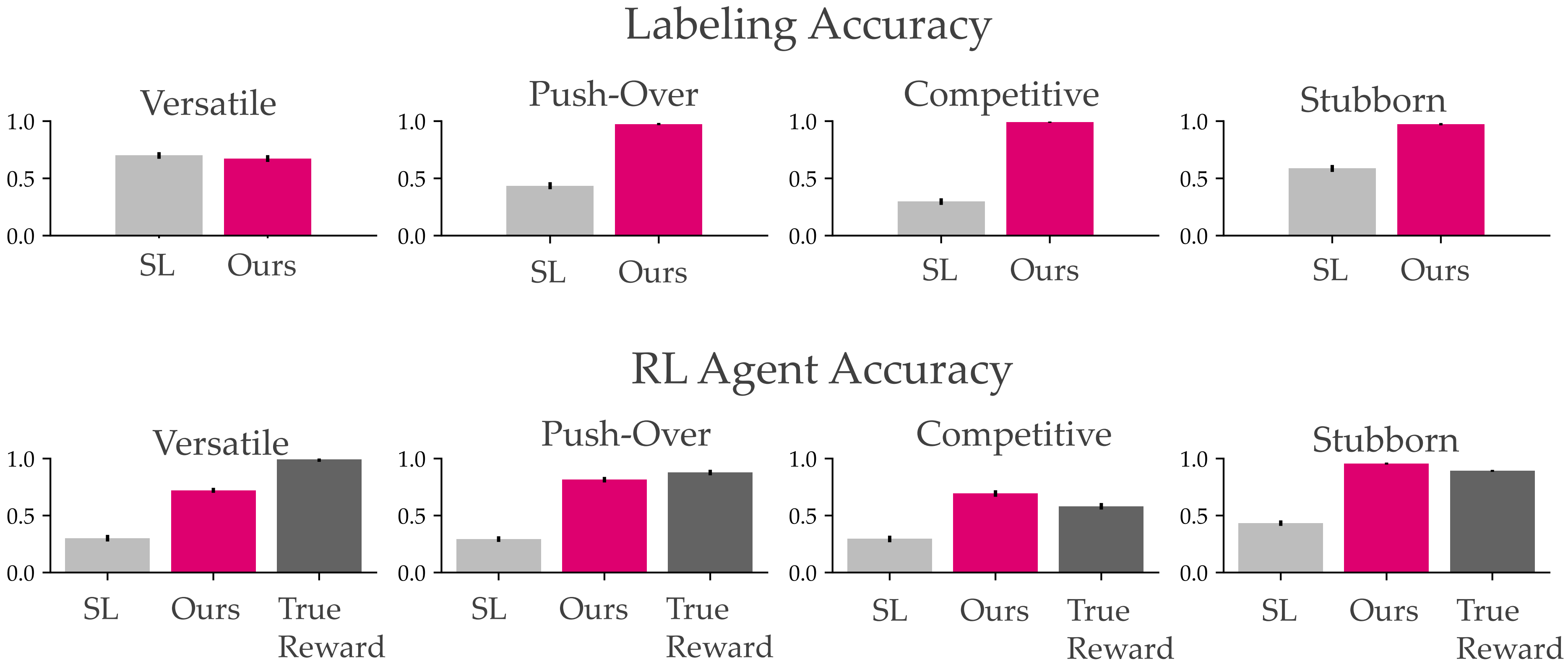
Automated Metrics (Ground Truth Rewards)

- **Versatile**: *Alice* does not suggest the same proposal more than once
- **Push-Over**: *Alice* gets less points than *Bob*
- **Competitive**: *Alice* gets more points than *Bob*
- **Stubborn**: *Alice* repeatedly suggests the same proposal

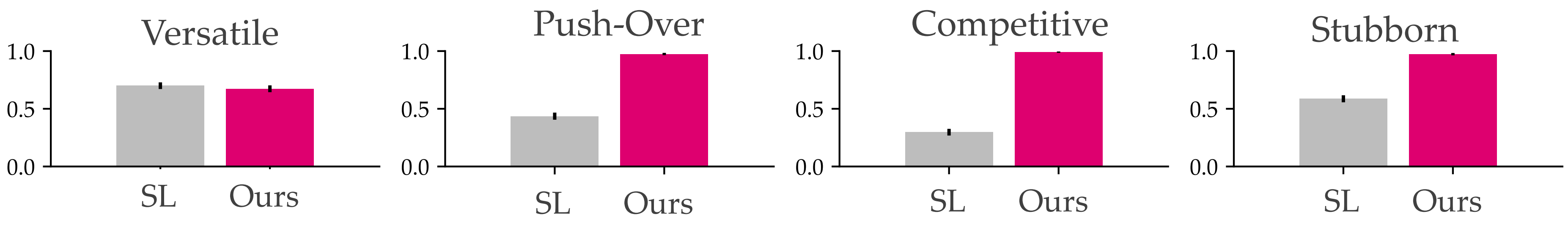
Baseline:

- A supervised learning (SL) model trained to predict reward signals using the same examples given to the LLM in our framework


Labeling Accuracy
Versatile
Push-Over
Competitive
Stubborn
1.0
0.5
0.0
SL
Ours


Labeling Accuracy
Versatile
Push-Over
Competitive
Stubborn
1.0
0.5
0.0
SL
Ours
RL Agent Accuracy
Versatile
Push-Over
Competitive
Stubborn
1.0
0.5
0.0
SL
Ours
True Reward

Labeling Accuracy

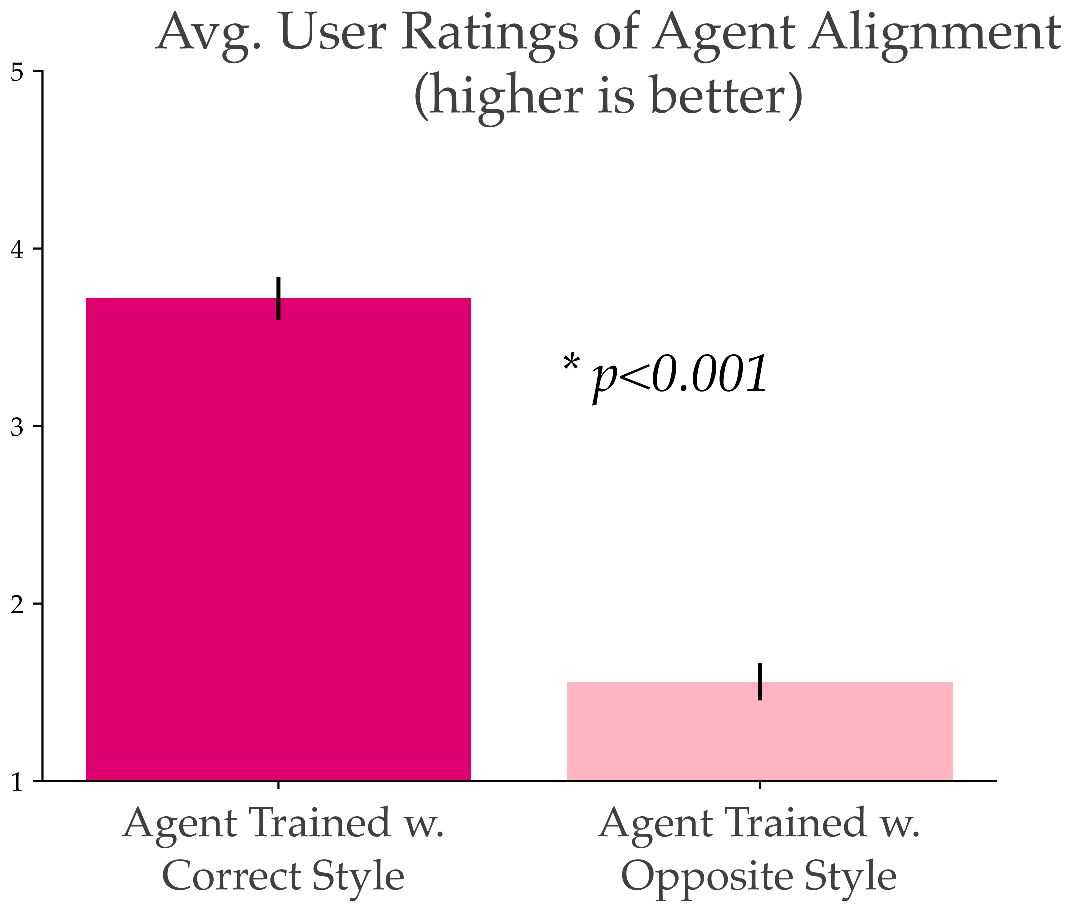


RL Agent Accuracy

We outperform SL by avg. of 46%

We underperform True Reward by avg. of 4%

We can use an LLM as a proxy reward to train objective-aligned agents

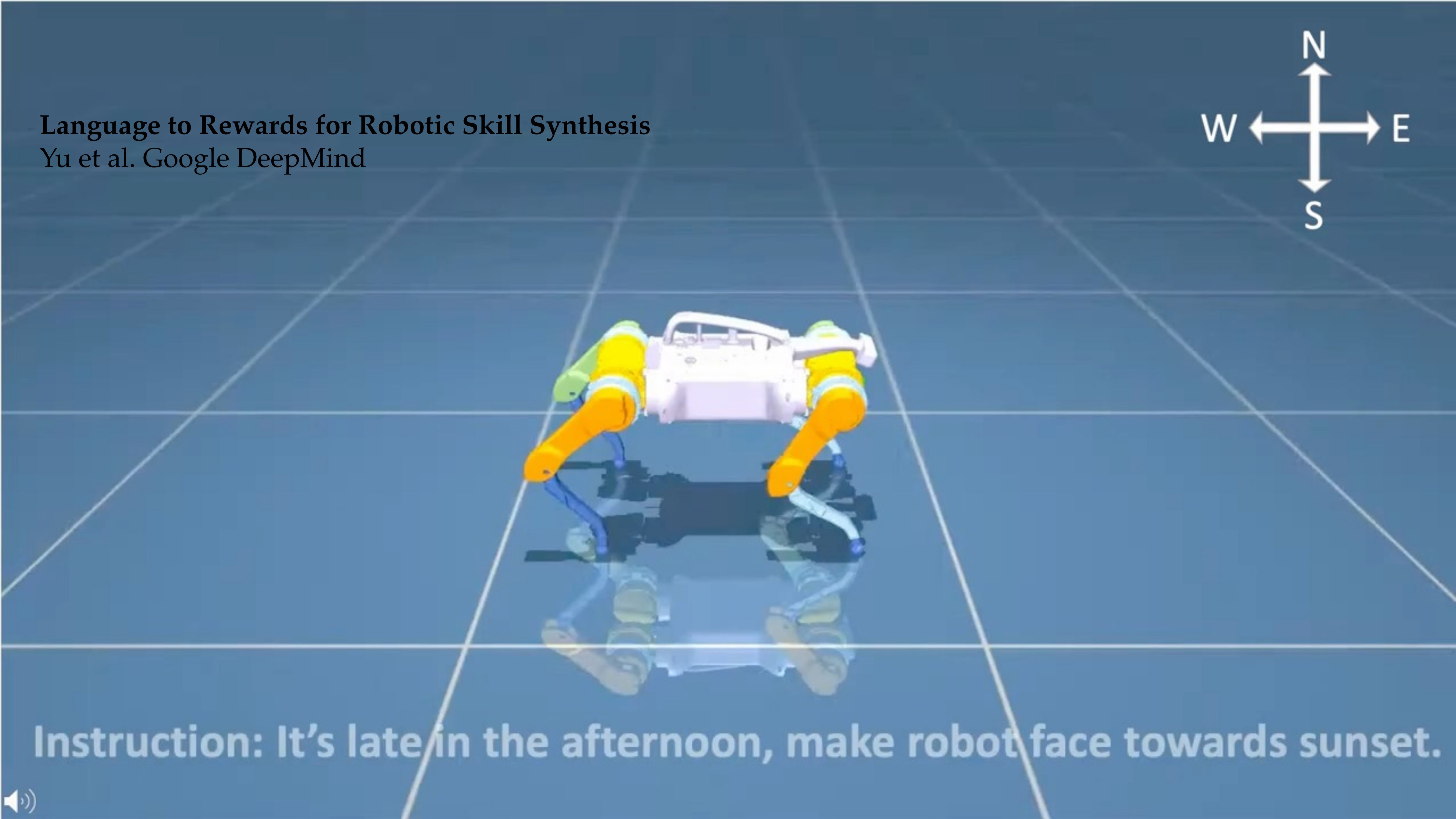


N=10

Examples of styles our users chose:
Polite, Push-Over, Considerate, Compromising, Ambitious

Humans find our agents more aligned
than an agent trained with a different objective.

Language to Rewards for Robotic Skill Synthesis
Yu et al. Google DeepMind

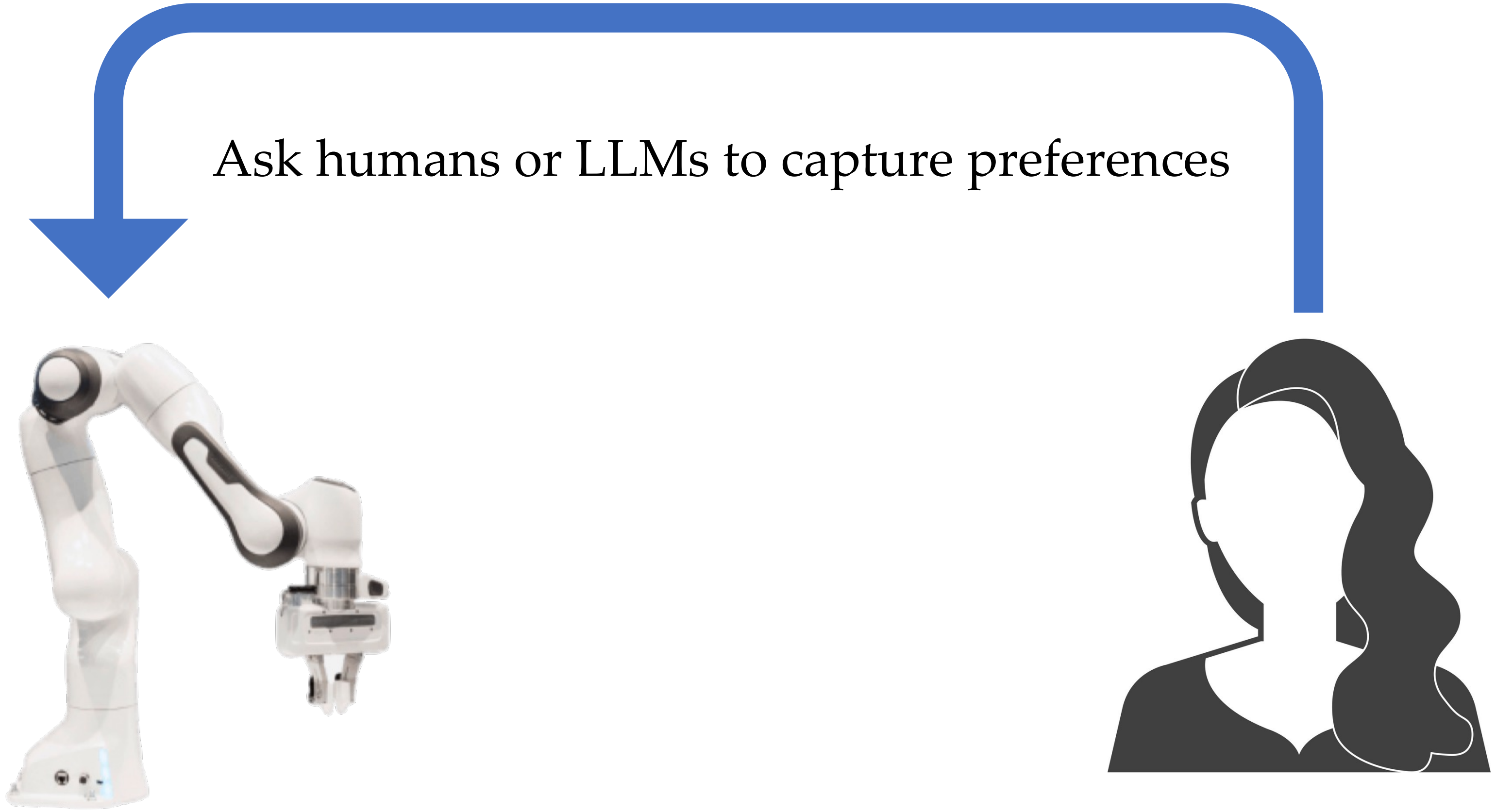


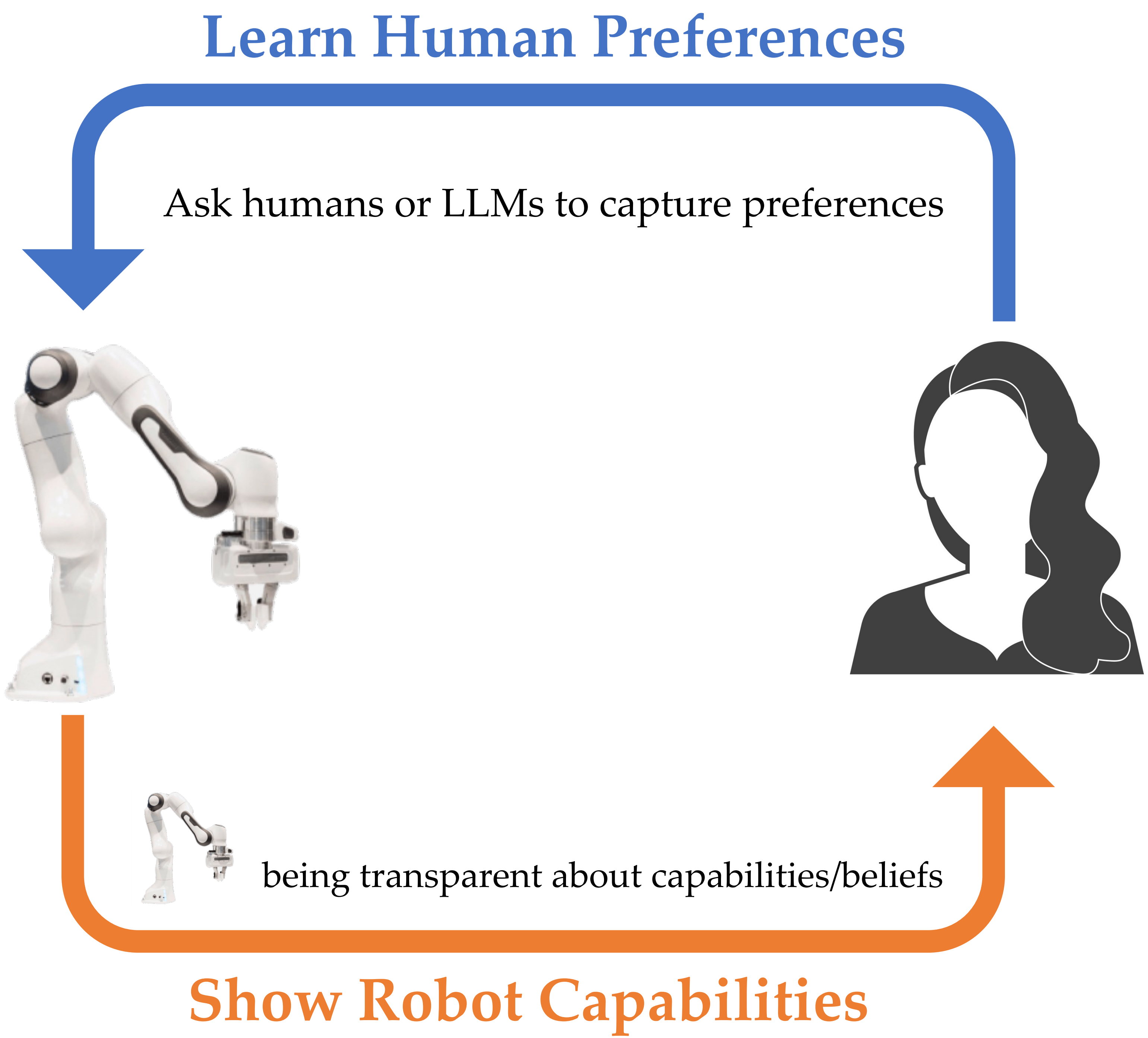
Key Takeaway 1

We can learn human preference reward functions by

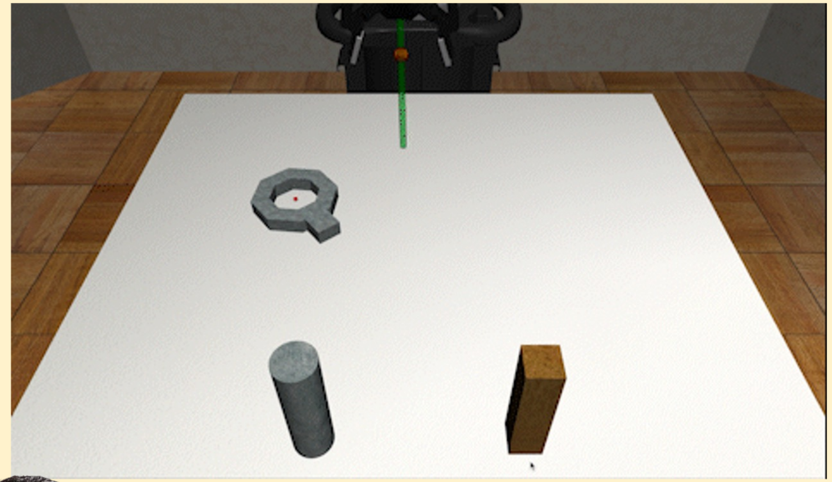
1) Actively querying for informative human feedback
2) leveraging the knowledge of large language models.

Learn Human Preferences

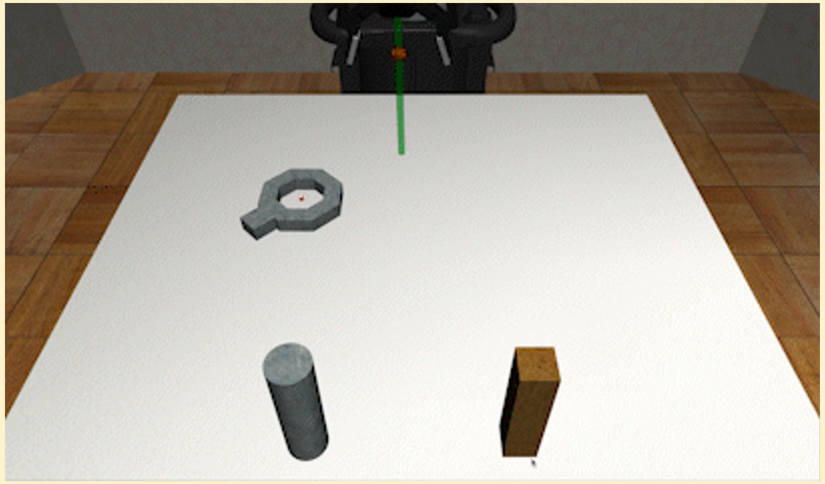



Learn Human Preferences
Ask humans or LLMs to capture preferences
being transparent about capabilities/beliefs
Show Robot Capabilities

What happens when multiple people teach?



+



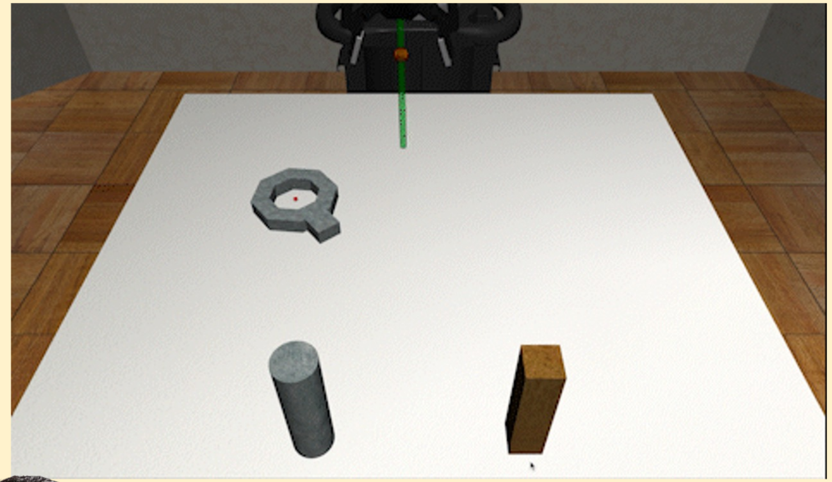
Kanishk

Sidd

14% rate of success

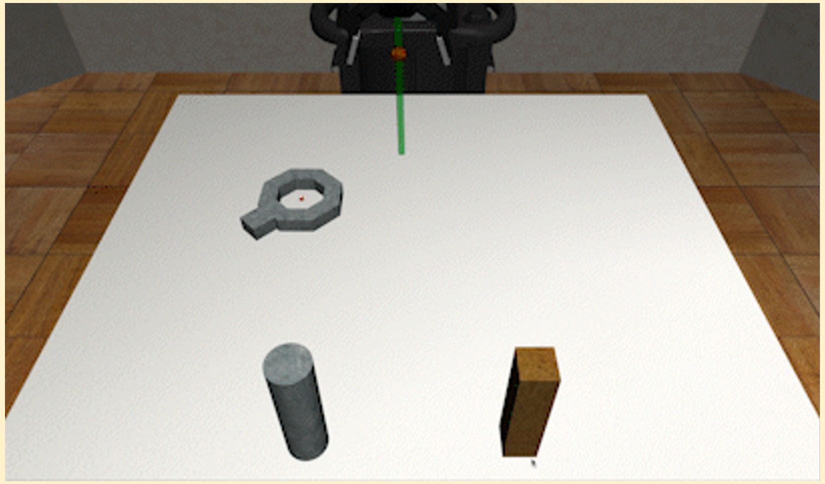
Kanishk Only

What happens when multiple people teach?



Kanishk

+



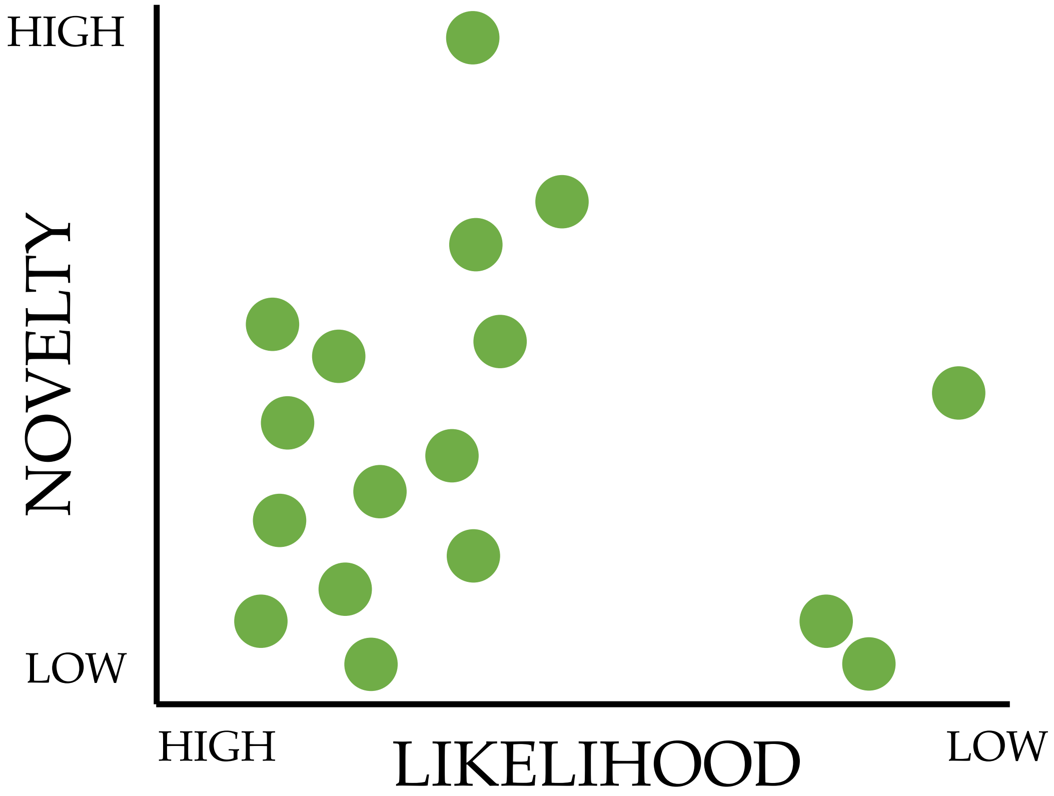
Sidd

14% rate of success — Kanishk Only

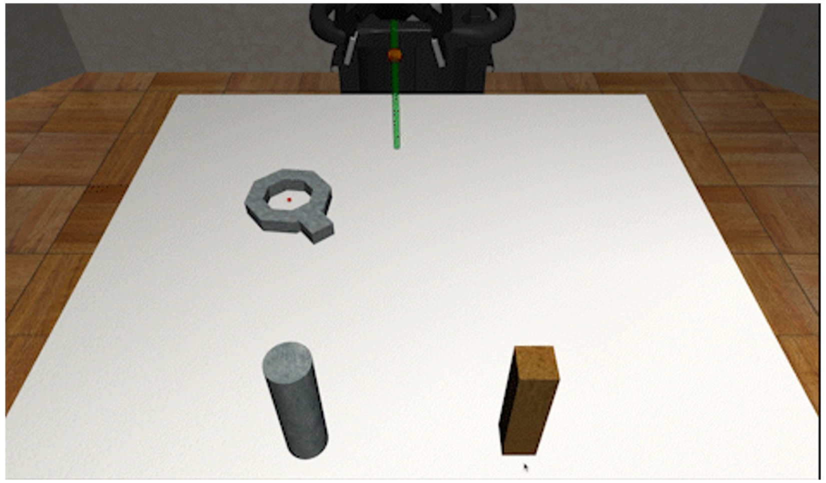
7% rate of success — Kanishk + Sidd

What happens when multiple people teach?

A Tale of Two Measures: Novelty and Likelihood

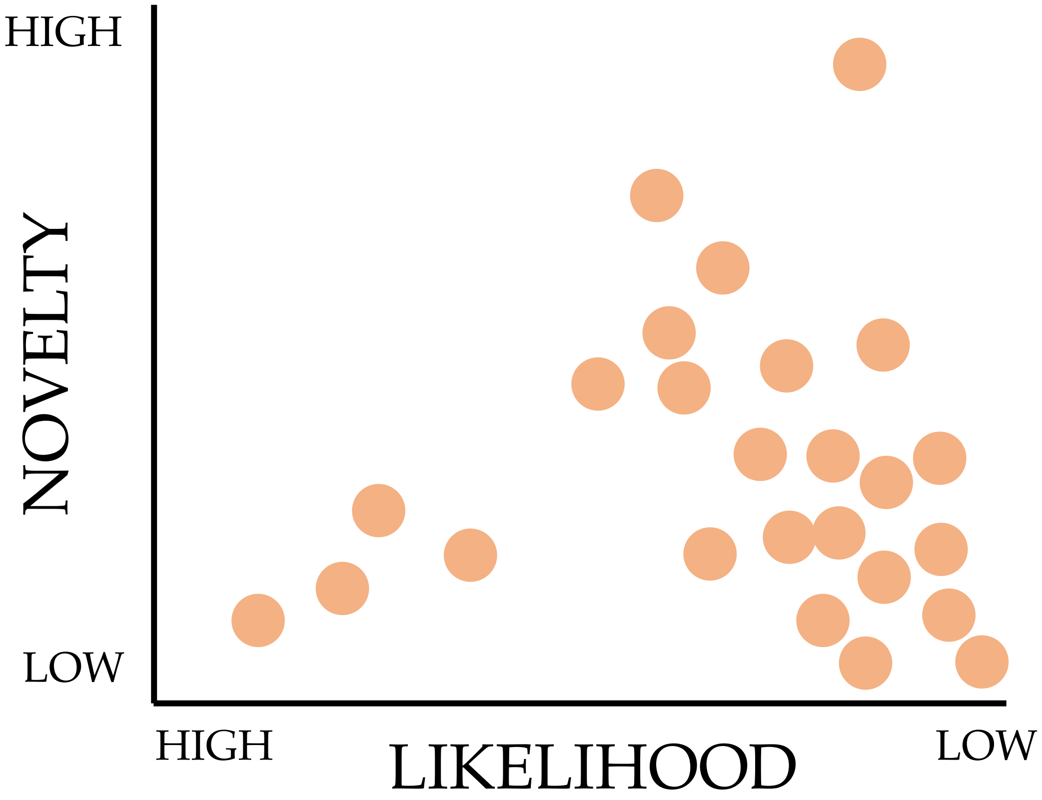


Kanishk's Data

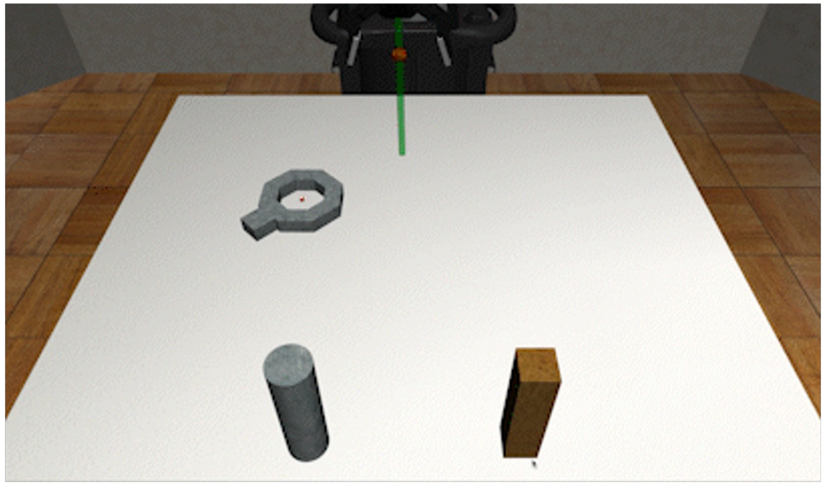


Eliciting Compatible Demonstrations for Multi-Human Imitation Learning
Gandhi, Karamcheti, Liao, Sadigh
CoRL 2022

A Tale of Two Measures: Novelty and Likelihood



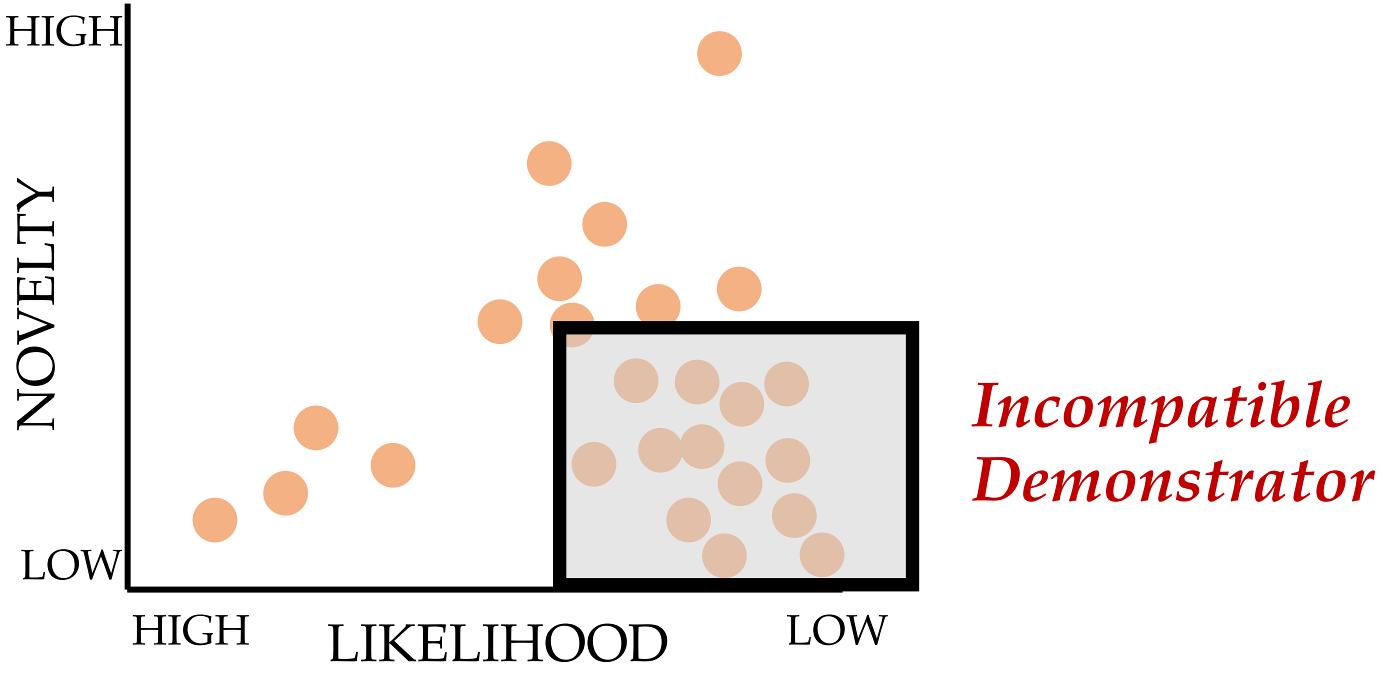
Sidd's Data



Eliciting Compatible Demonstrations for Multi-Human Imitation Learning
Gandhi, Karamcheti, Liao, Sadigh
CoRL 2022

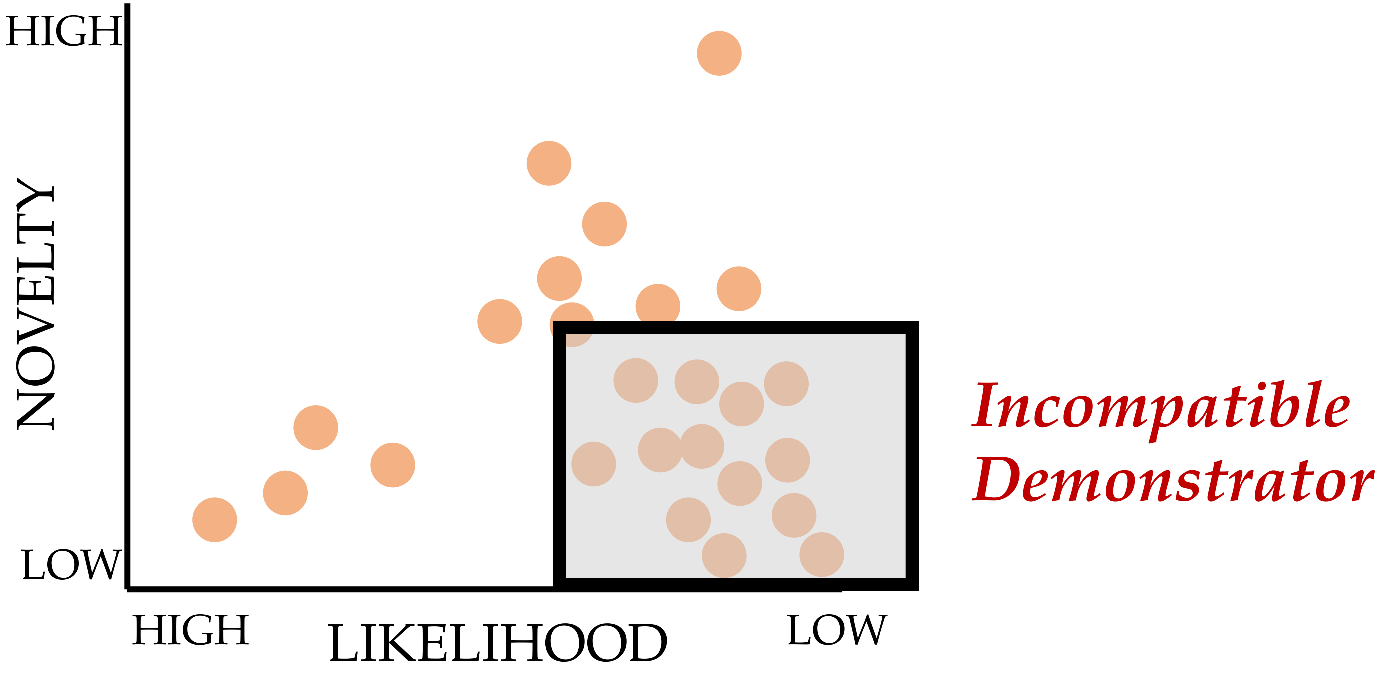
Filtering demonstrations based on compatibility

Operator	**Square Nut**	
	Naive	$\mathcal{M}$-Filtered
Base Operator	38.7 (2.1)	-



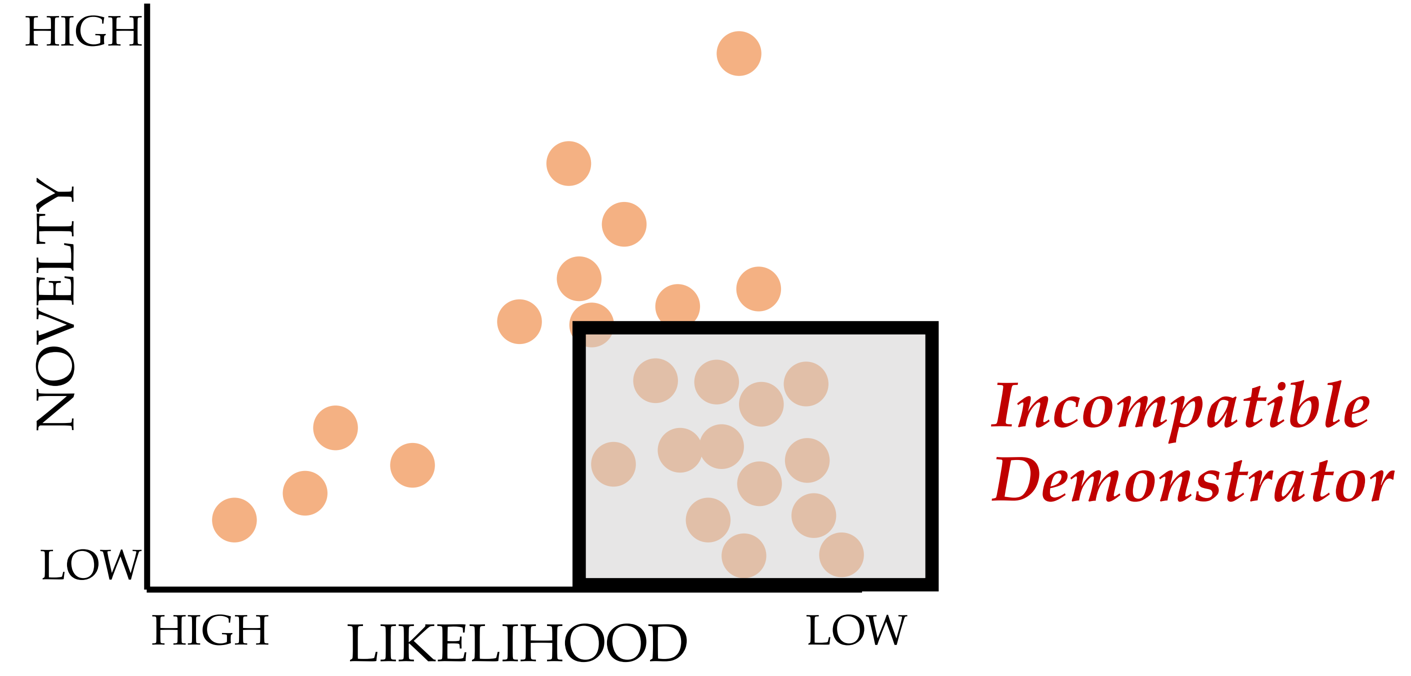
Filtering demonstrations based on compatibility

Operator	Square Nut	
	Naive	$\mathcal{M}$-Filtered
Base Operator	38.7 (2.1)	-
Operator 1	54.3 (1.5)	61.0 (4.4)



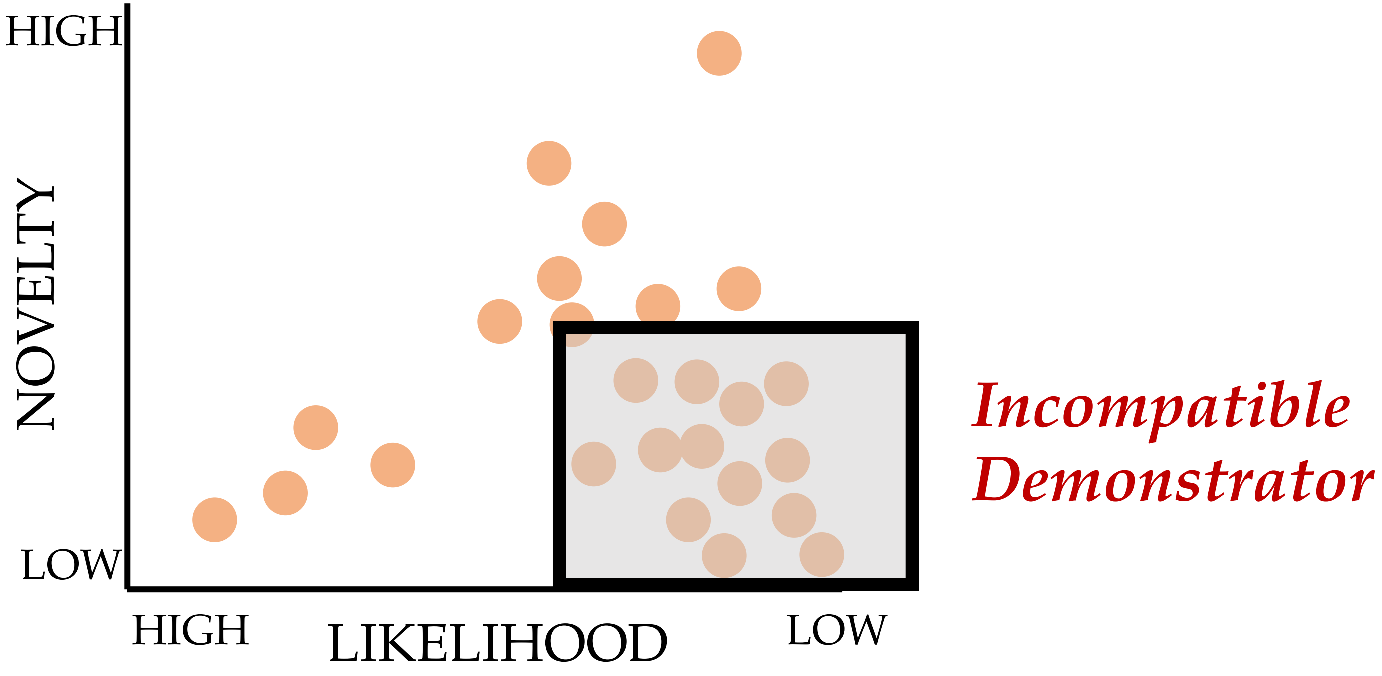
Filtering demonstrations based on compatibility

Operator	Square Nut		Round Nut		Hammer Placement	
	Naive	$\mathcal{M}$-Filtered	Naive	$\mathcal{M}$-Filtered	Naive	$\mathcal{M}$-Filtered
Base Operator	38.7 (2.1)	-	13.3 (2.3)	-	24.7 (6.1)	-
Operator 1	54.3 (1.5)	61.0 (4.4)	26.7 (11.7)	32.0 (12.2)	38.0 (2.0)	39.7 (4.6)



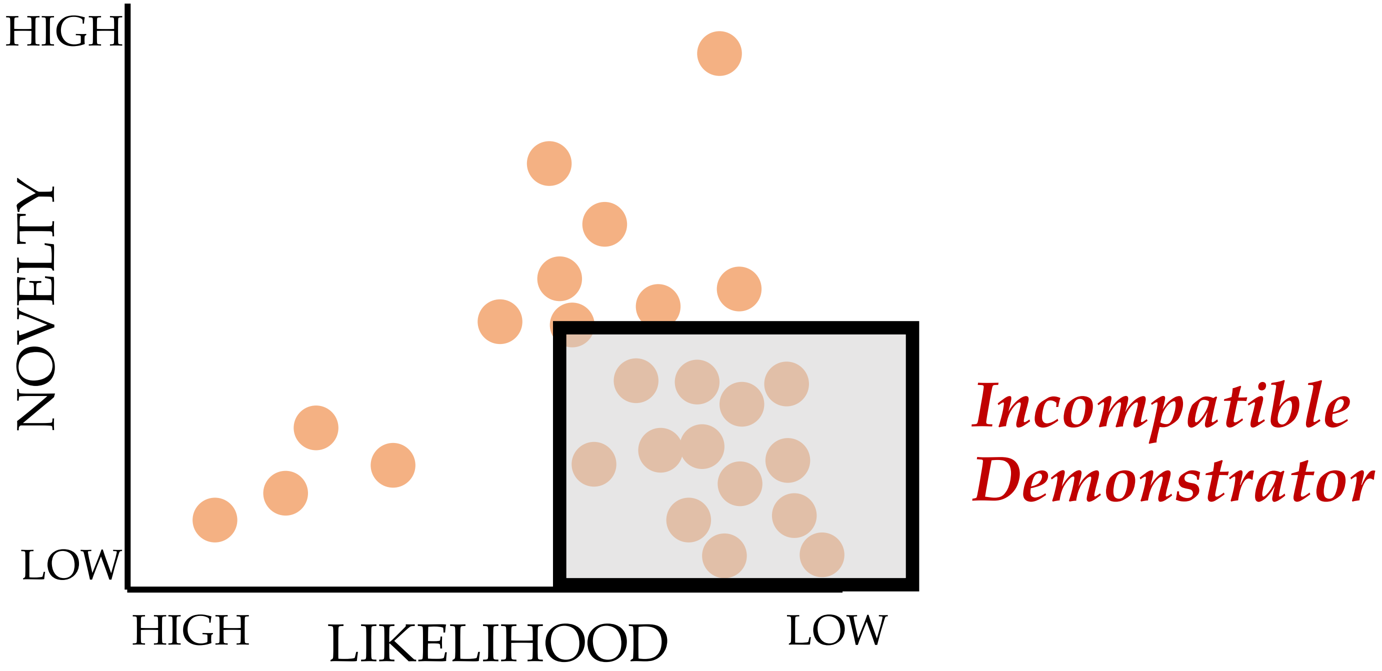
Filtering demonstrations based on compatibility

Operator	Square Nut		Round Nut		Hammer Placement	
	Naive	$\mathcal{M}$-Filtered	Naive	$\mathcal{M}$-Filtered	Naive	$\mathcal{M}$-Filtered
Base Operator	38.7 (2.1)	-	13.3 (2.3)	-	24.7 (6.1)	-
Operator 1	54.3 (1.5)	61.0 (4.4)	26.7 (11.7)	32.0 (12.2)	38.0 (2.0)	39.7 (4.6)
Operator 2	40.3 (5.1)	42.0 (2.0)	22.0 (7.2)	26.7 (5.0)	33.3 (3.1)	32.7 (6.4)
Operator 3	37.3 (2.1)	42.7 (0.6)	17.3 (4.6)	18.0 (13.9)	8.0 (0.0)	12.0 (0.0)
Operator 4	27.3 (3.5)	37.3 (2.1)	7.3 (4.6)	13.3 (1.2)	4.0 (0.0)	4.0 (0.0)

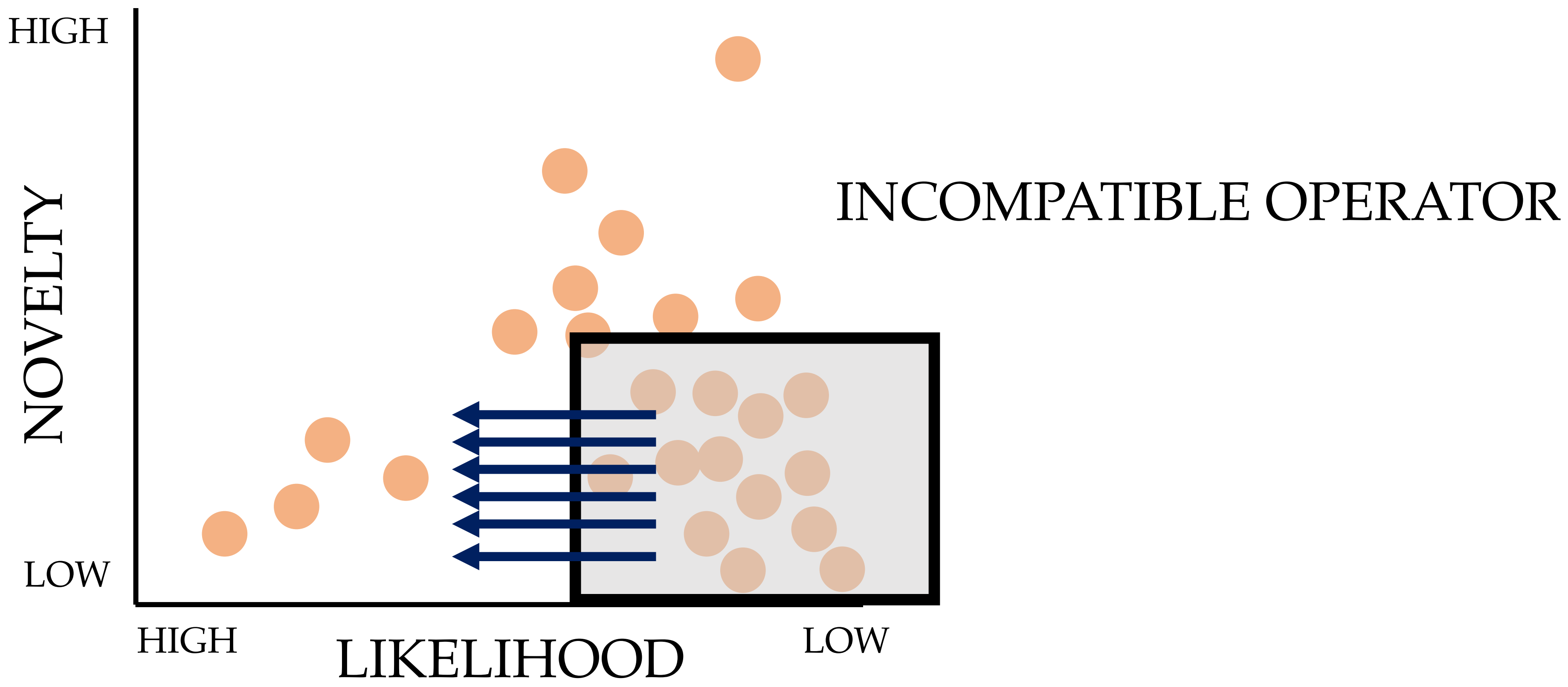


Filtering demonstrations based on compatibility

Operator	Square Nut		Round Nut		Hammer Placement	
	Naive	$\mathcal{M}$-Filtered	Naive	$\mathcal{M}$-Filtered	Naive	$\mathcal{M}$-Filtered
Base Operator	38.7 (2.1)	-	13.3 (2.3)	-	24.7 (6.1)	-
Operator 1	54.3 (1.5)	61.0 (4.4)	26.7 (11.7)	32.0 (12.2)	38.0 (2.0)	39.7 (4.6)
Operator 2	40.3 (5.1)	42.0 (2.0)	22.0 (7.2)	26.7 (5.0)	33.3 (3.1)	32.7 (6.4)
Operator 3	37.3 (2.1)	42.7 (0.6)	17.3 (4.6)	18.0 (13.9)	8.0 (0.0)	12.0 (0.0)
Operator 4	27.3 (3.5)	37.3 (2.1)	7.3 (4.6)	13.3 (1.2)	4.0 (0.0)	4.0 (0.0)

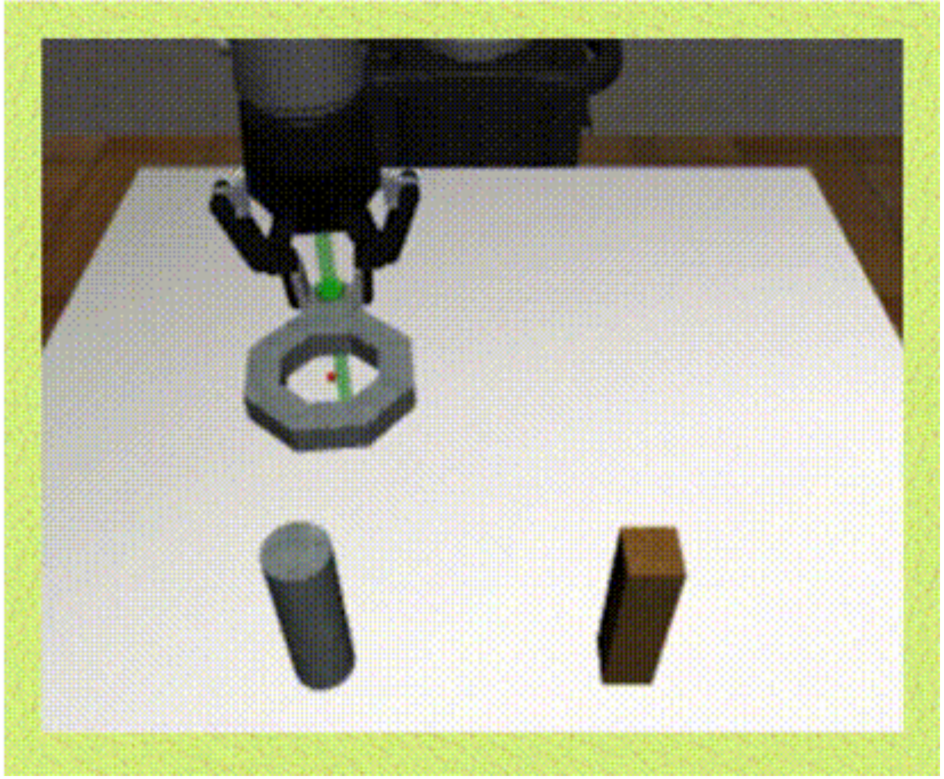


Guiding demonstrations based on compatibility



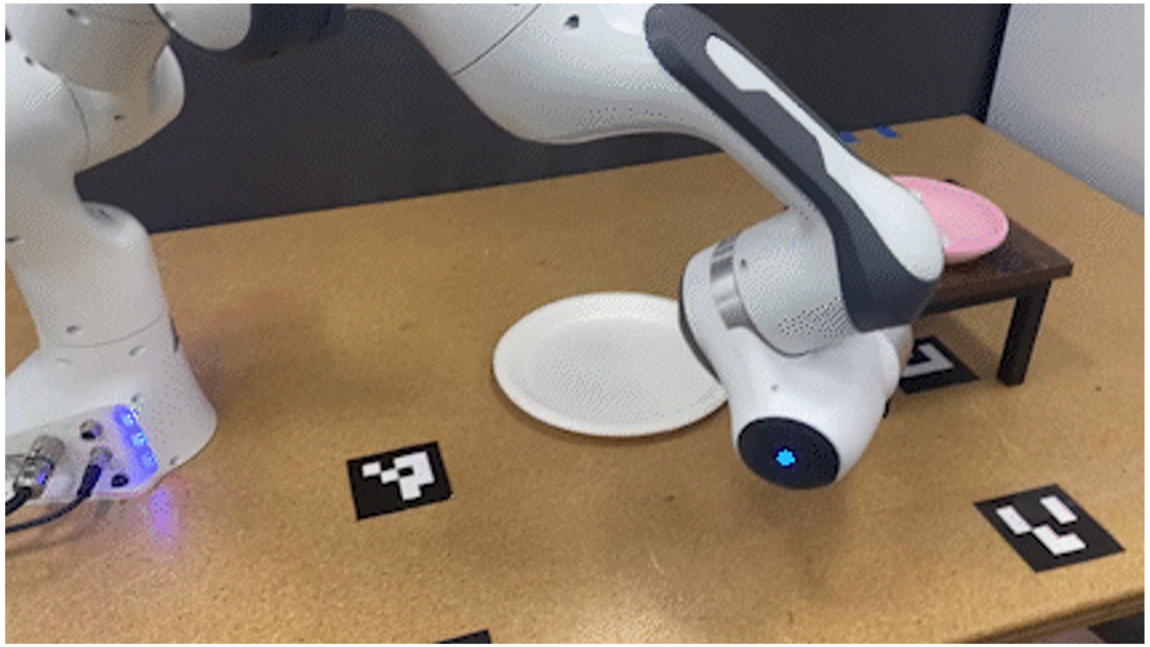
Active Elicitation Interface

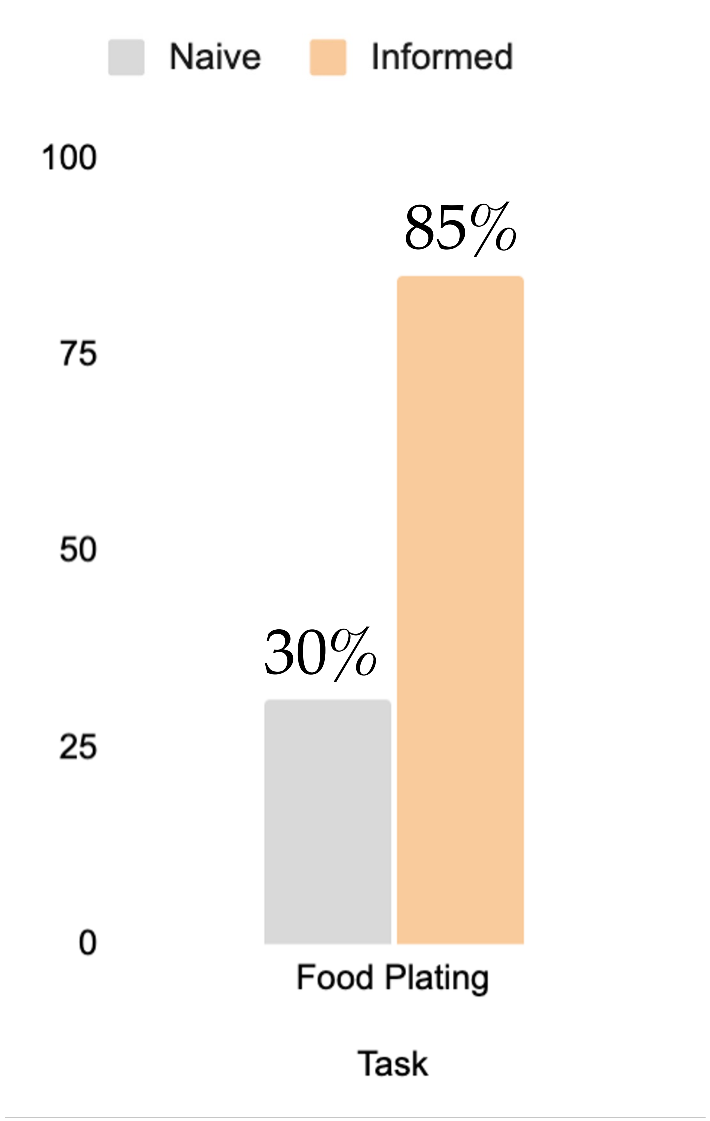
Interactively show the demonstrator if the actions are compatible or not



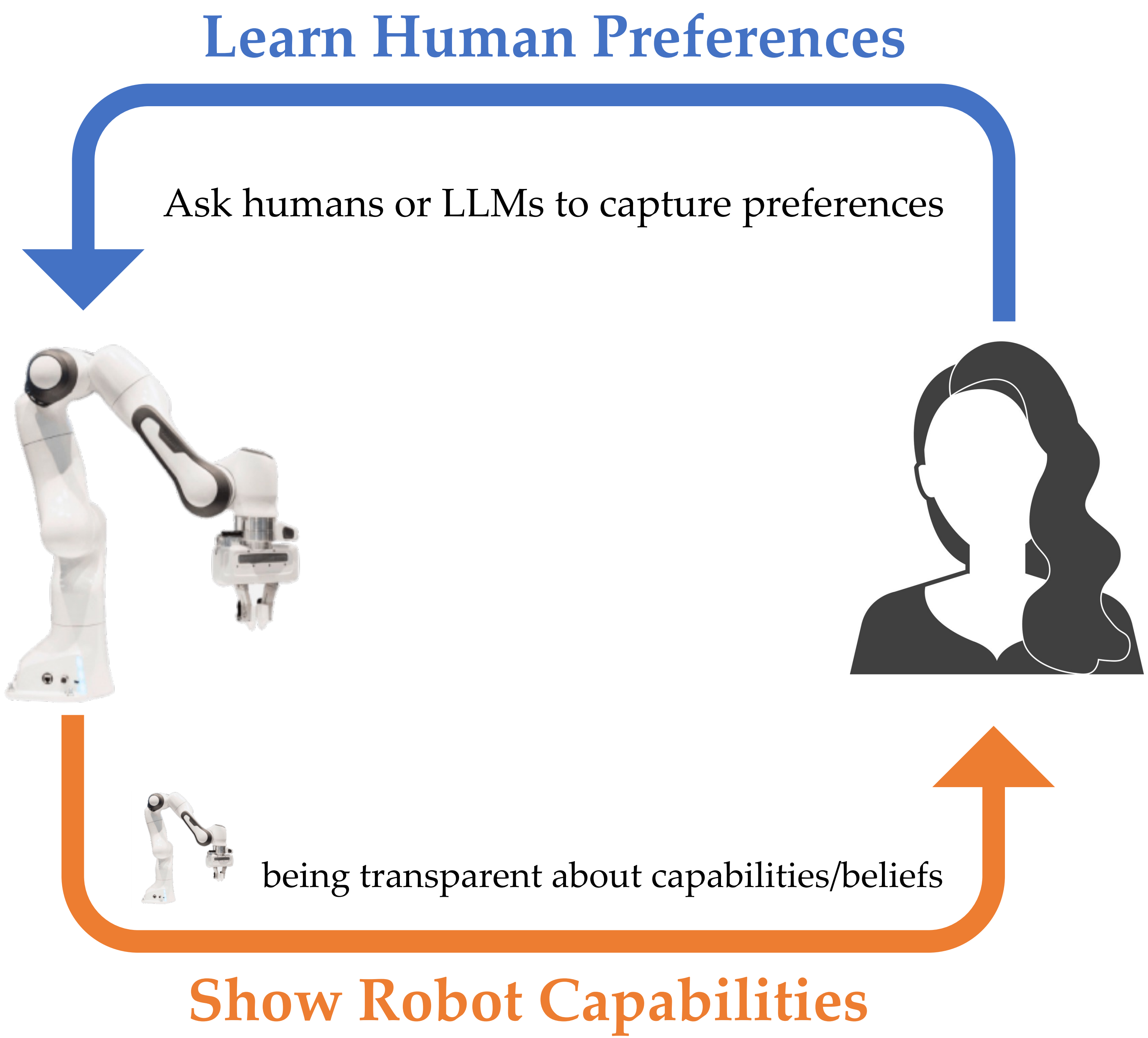
Task	Base	Naïve	Naïve + Filtered	Informed
Round Nut	13.3 (2.3)	9.6 (4.6)	9.7 (4.2)	15.7 (6.0)
Hammer Placement	24.7 (6.1)	20.8 (15.7)	22.0 (15.5)	31.8 (16.3)
[Real] Food Plating	60.0	30.0 (17.3)	-	85.0 (9.6)

How do policies from informed demonstrators perform?





Eliciting Compatible Demonstrations for Multi-Human Imitation Learning
Gandhi, Karamcheti, Liao, Sadigh
CoRL 2022


Learn Human Preferences
Ask humans or LLMs to capture preferences
being transparent about capabilities/beliefs
Show Robot Capabilities

Key Takeaway 1

We can learn human preference reward functions by

1) Actively querying for informative human feedback
2) leveraging the knowledge of large language models.

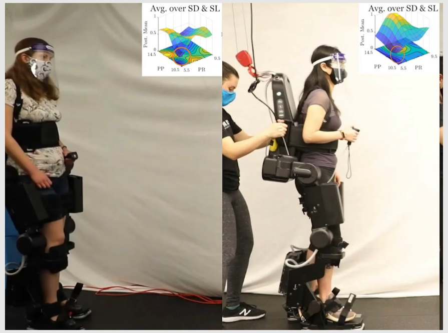
Key Takeaway 1

We can learn human preference reward functions by

1) Actively querying for informative human feedback
2) leveraging the knowledge of large language models.

We can ask humans to do more than answering question…
Transparent robots can guide the human to provide compatible demonstrations.

Learning Human Preferences

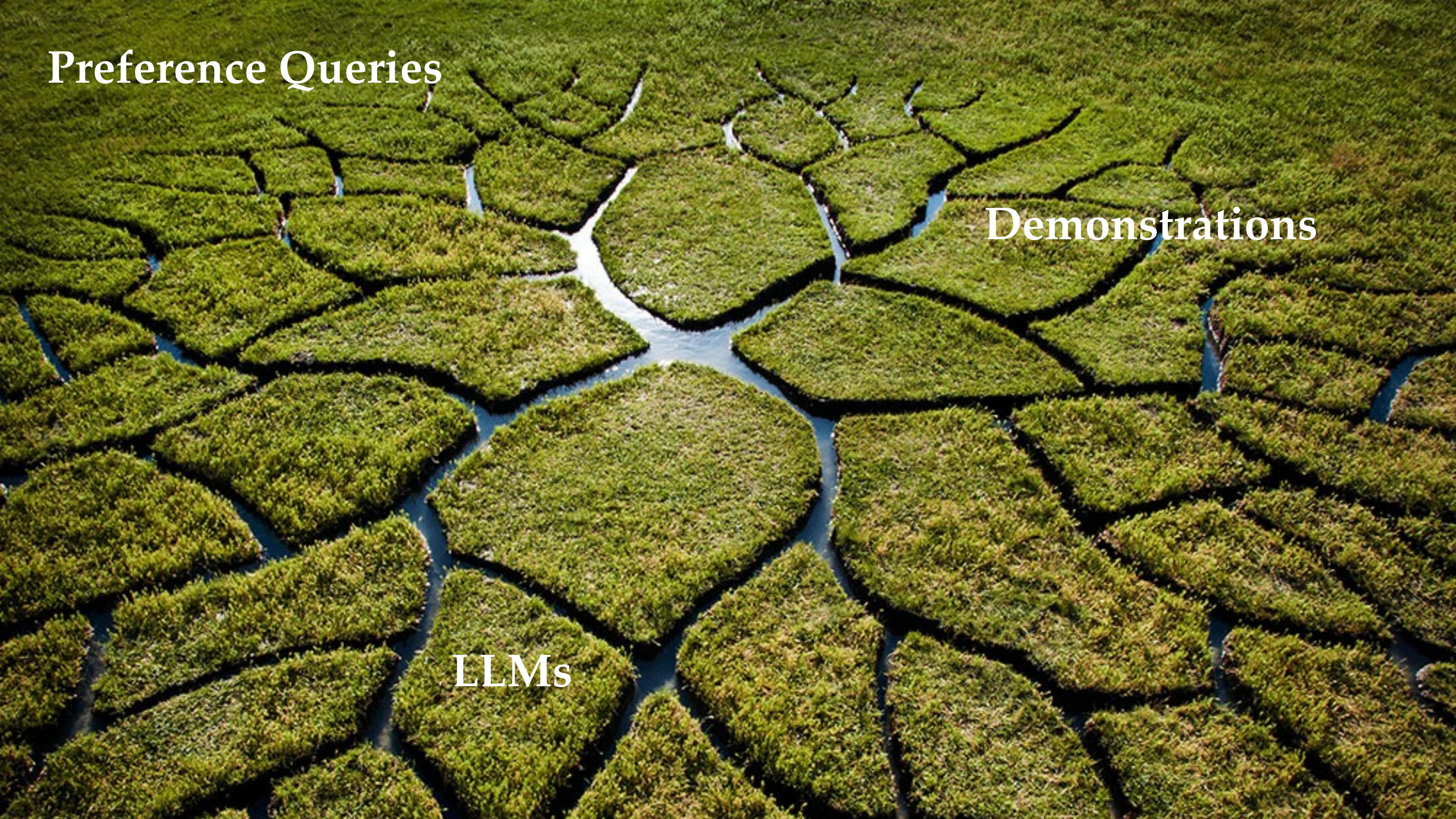


Biyik et al. IJRR 21
Kwon et al. ICLR 23
Gandhi et al. CoRL 22

Foundation Models for Robotics

Voltron

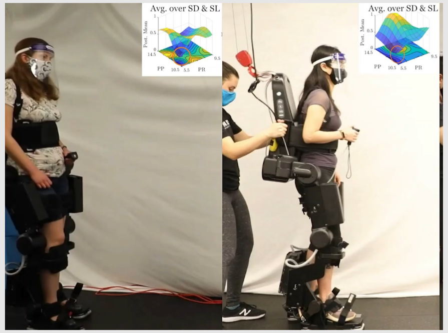
Karamcheti et al. RSS23
Mirchandani et al. CoRL23


Preference Queries
Demonstrations
LLMs

Large Language Models are now a thing…

What does that mean for robotics?

Learning Human Preferences



Biyik et al. IJRR 21
Kwon et al. ICLR 23
Gandhi et al. CoRL 22

Foundation Models for Robotics

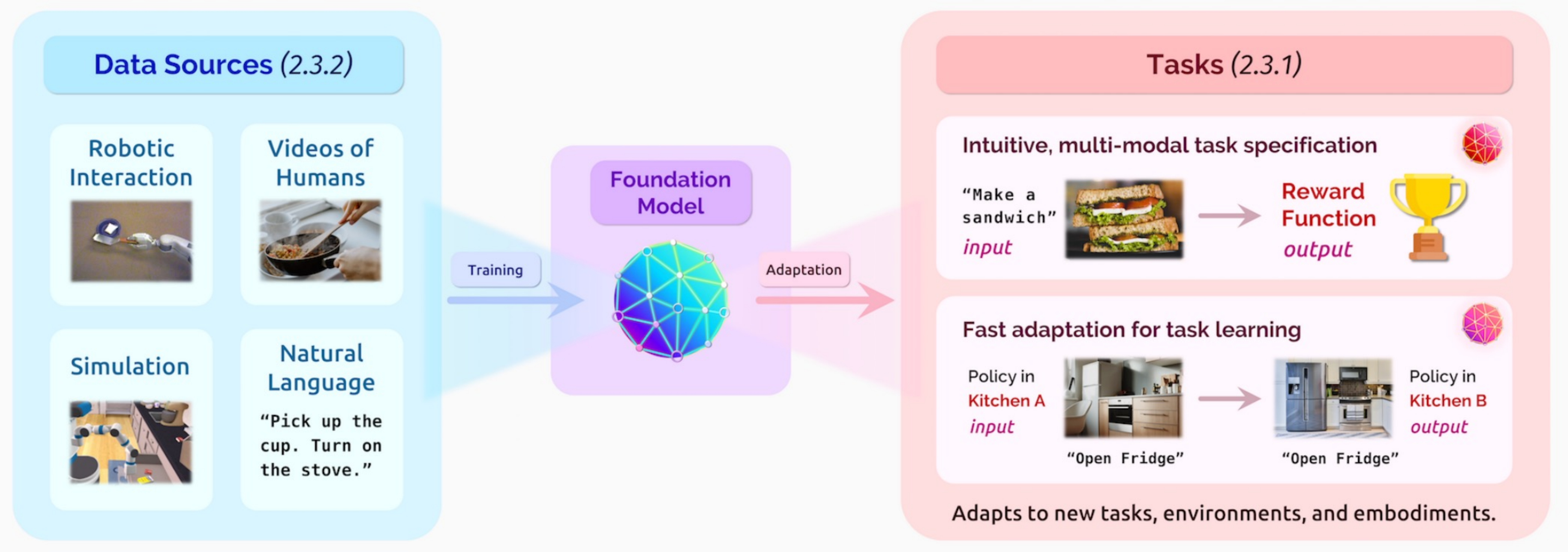
Voltron

Karamcheti et al. RSS23
Mirchandani et al. CoRL23

Take 1: What does it take to build a robotics foundation model?

Instead of learning from preference queries or demonstrations,
can we tap into large offline datasets?

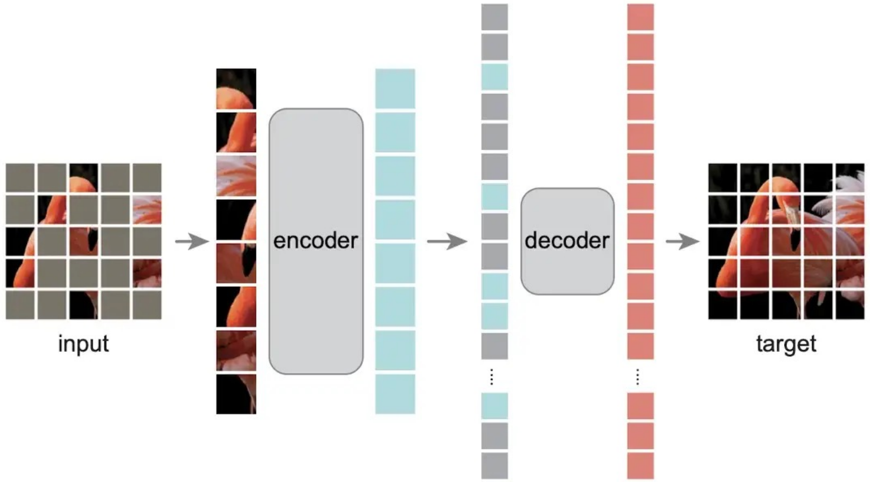
Robotics Foundation Models

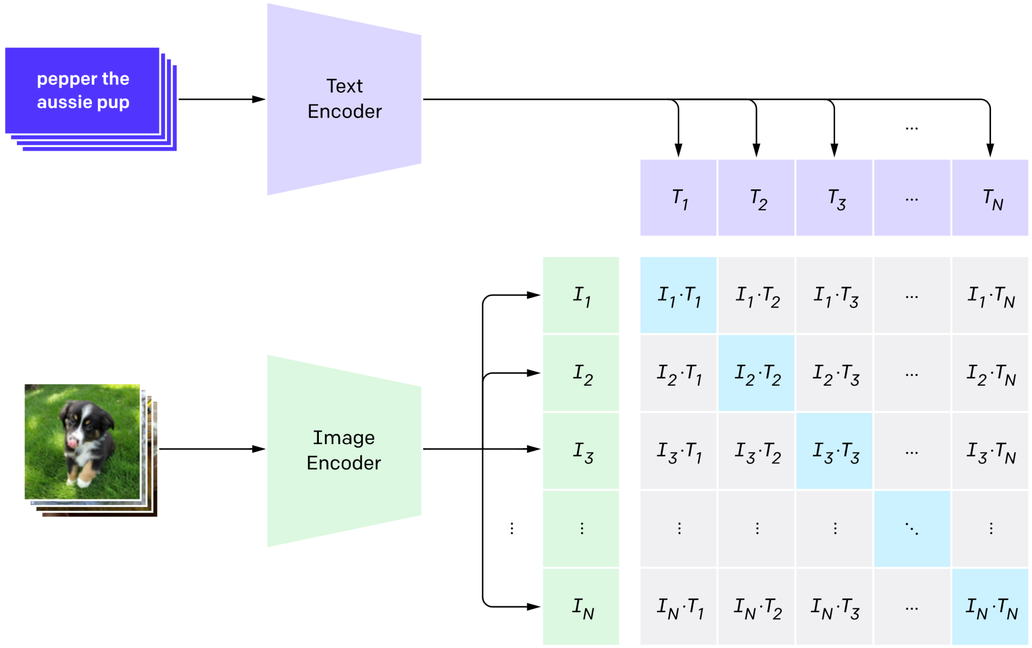


On the Opportunities and Risks of Foundation Models. Bommasani et al. 2021

Representation Learning for Robotics — Two Extremes

Existing work tends towards **specific visual representations** that are not flexible:





"Syntax" — Local/Spatial Features

MAE — Pixel Reconstruction

"learn patterns within an image"

"Semantics" — Generalizable Concepts

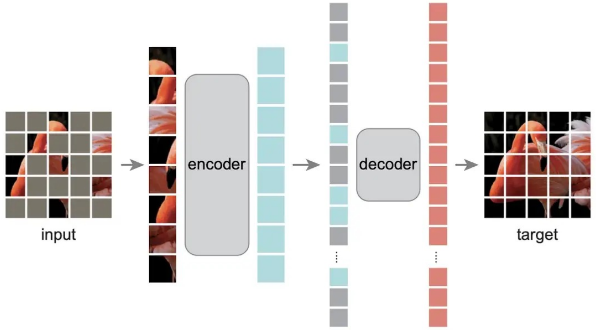
CLIP — Language Supervision

"learn concepts across images"

Key Idea: Use language supervision to shape representations!

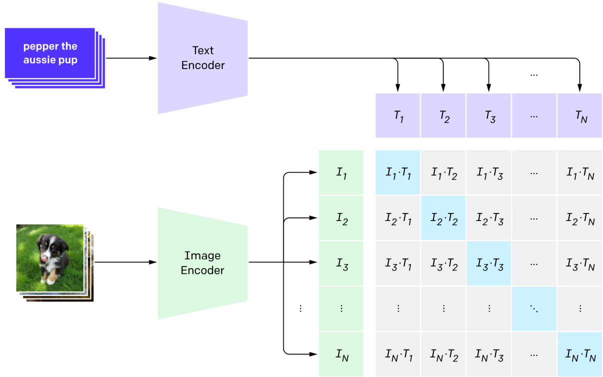
Best of Both Worlds — Bridging "Syntax" and "Semantics"

Key Idea: Use language supervision to shape representations!



Grounded Reconstruction

(conditioning on language)



"Syntax"

Reconstruction

(no language)

"Semantics"

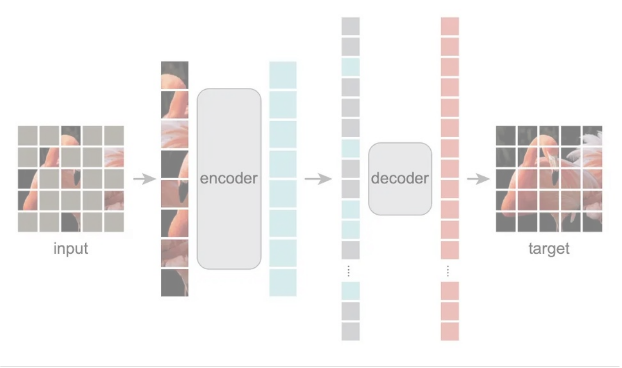
Captioning

(generating language)

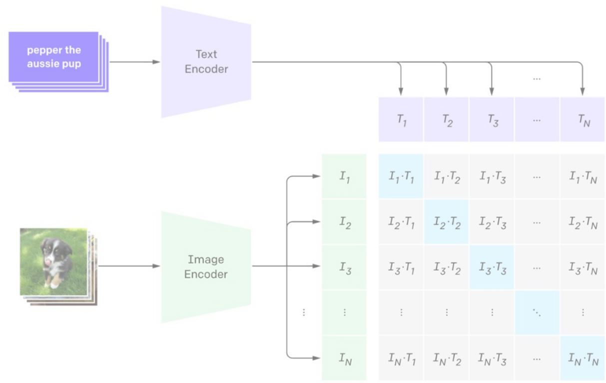
But… aren't we missing something!

Language-Driven Representation Learning

Key Idea: Use language supervision to shape representations!



"Syntax"

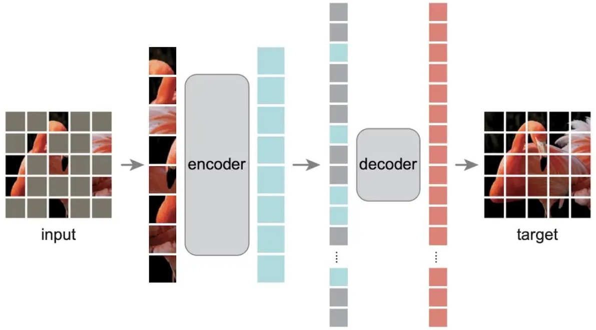


"Semantics"

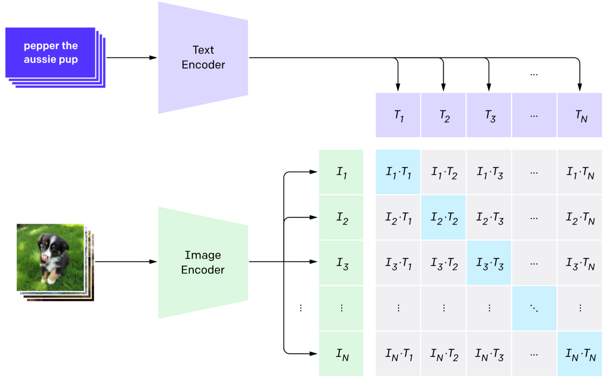
Modeling grounded, dynamic interactions atop syntax / semantics → "Pragmatics"

Voltron

Language-driven Representation Learning



Key Idea: Use language supervision to shape representations!



"Syntax"

"Semantics"

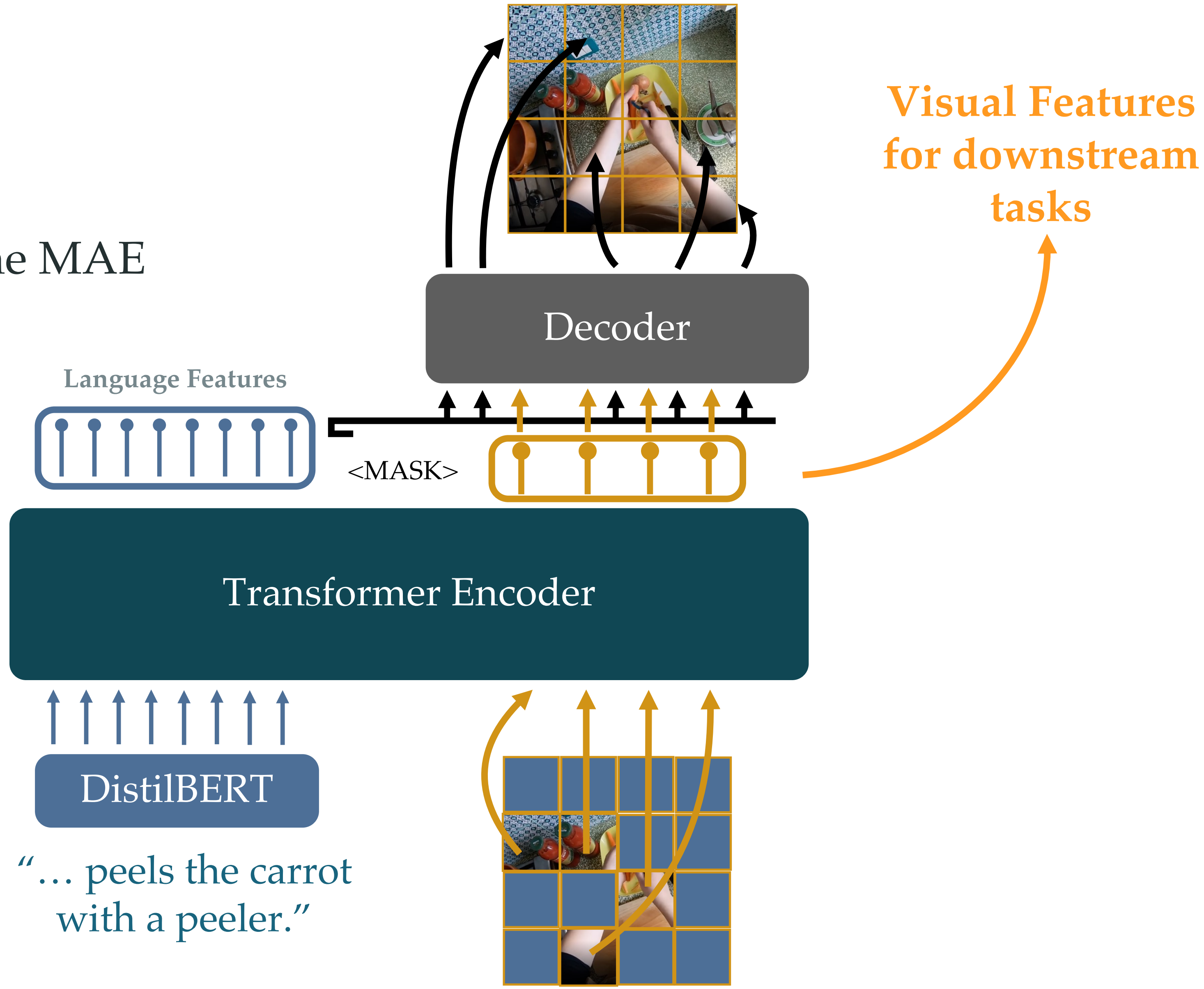
"Pragmatics"

Language-Driven Representation Learning for Robotics
S. Karamcheti, S. Nair, A. Chen, T. Kollar, C. Finn, D. Sadigh, P. Liang
Robotics: Science and Systems (RSS), 2023

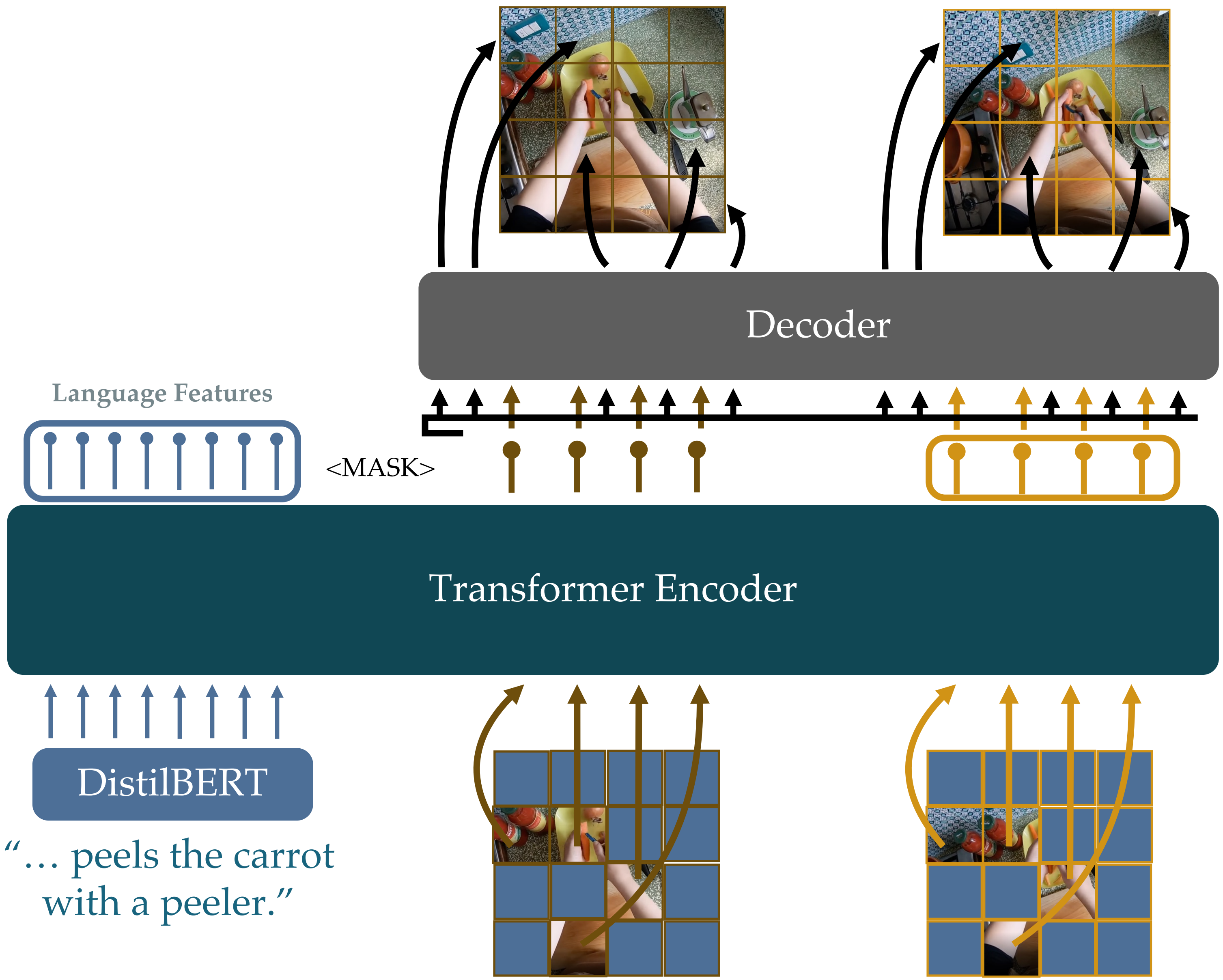
Sidd Karamcheti

Combining Syntax and Semantics

Enrich the base model by conditioning the MAE encoder on a ***language prefix.***

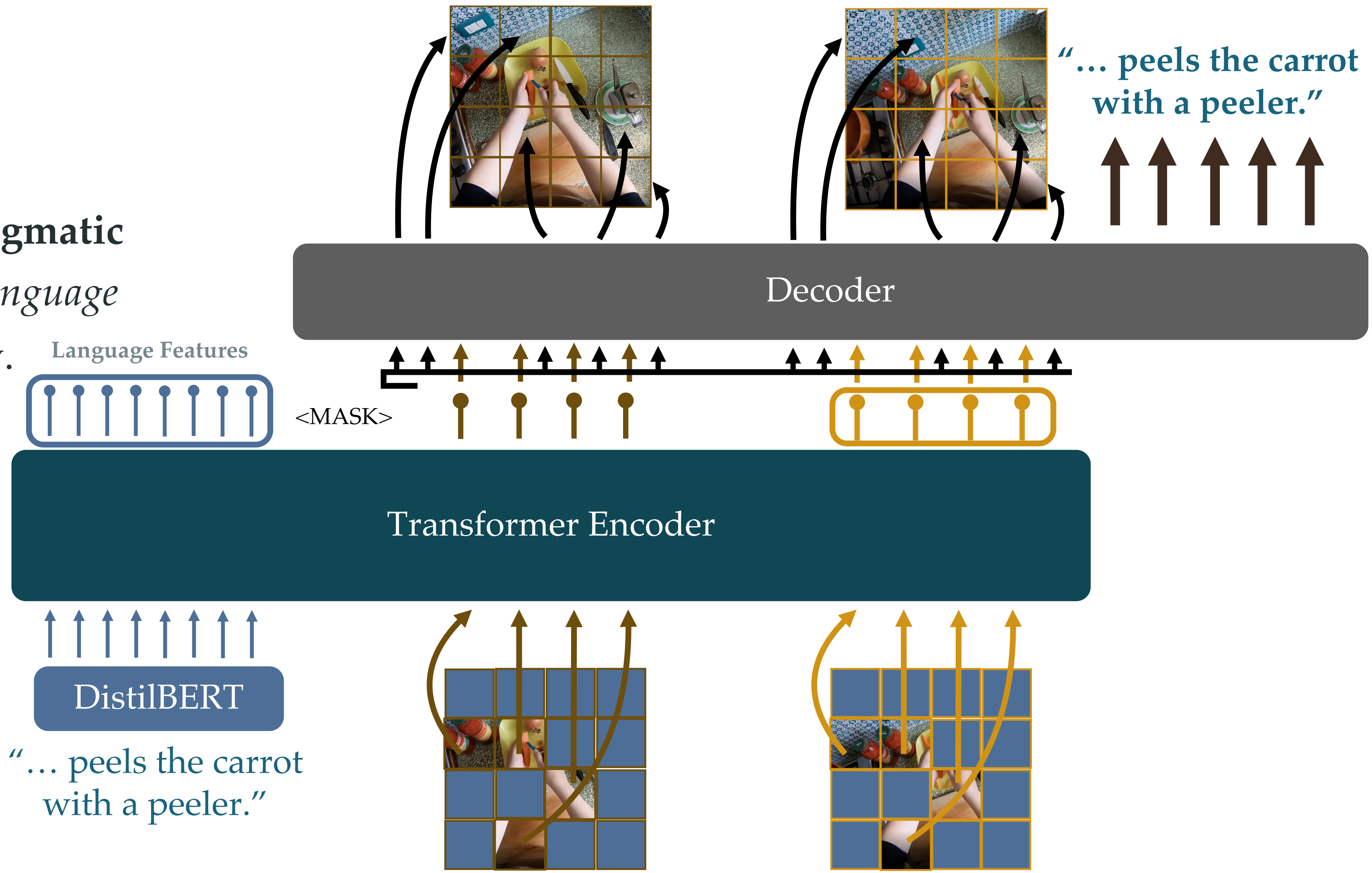


Adding **Pragmatics** (via Language Conditioning)



Adding **Pragmatics** (via Language Conditioning)

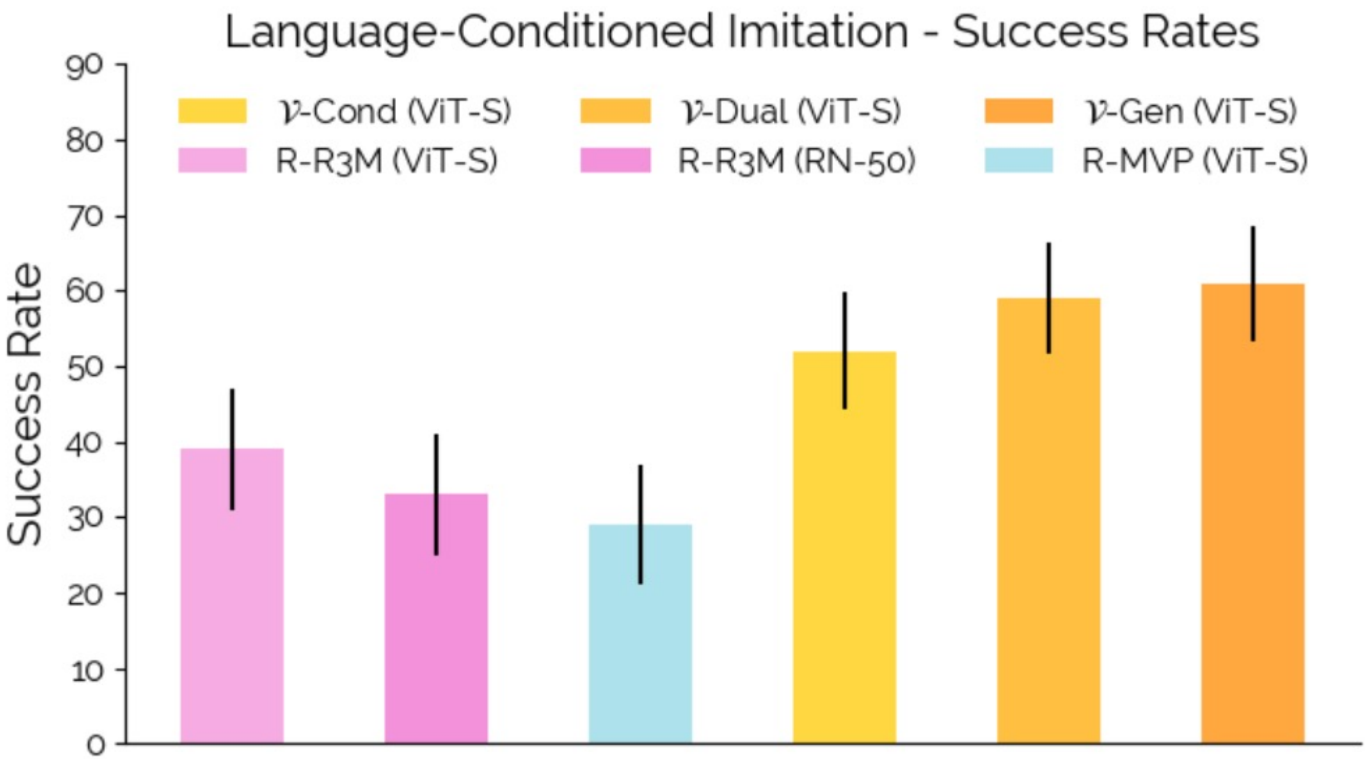
Boost **semantic** and **pragmatic** features by *generating language narrations*, given history.



Language-Conditioned Imitation Learning

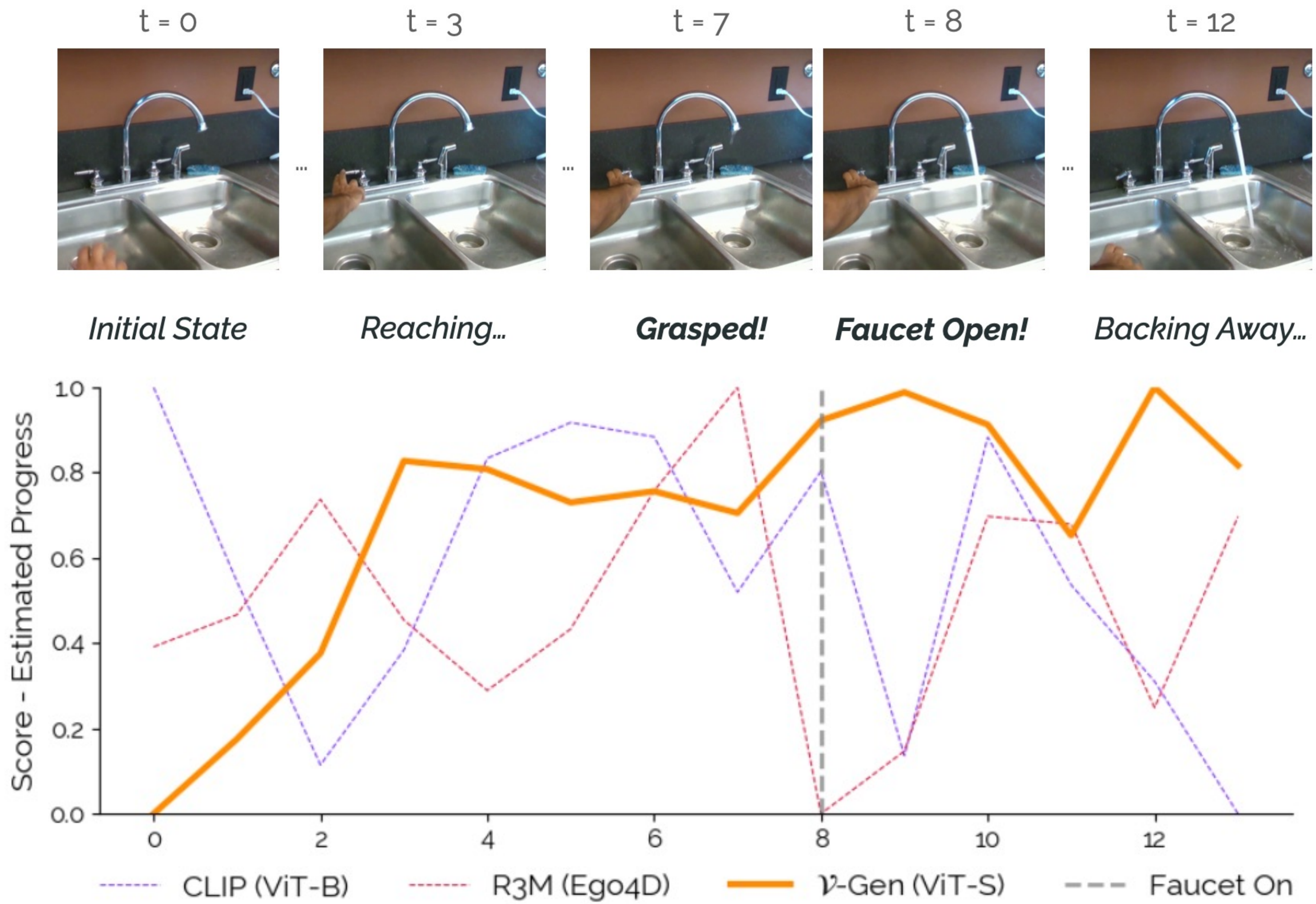
Study Desk Environment

- "Shut the drawer."
- "Throw the bag of chips away."
- "Discard the used coffee pods."
- "Put the blue mug on the purple plate."
- "Set the coffee on top of the yellow plate."

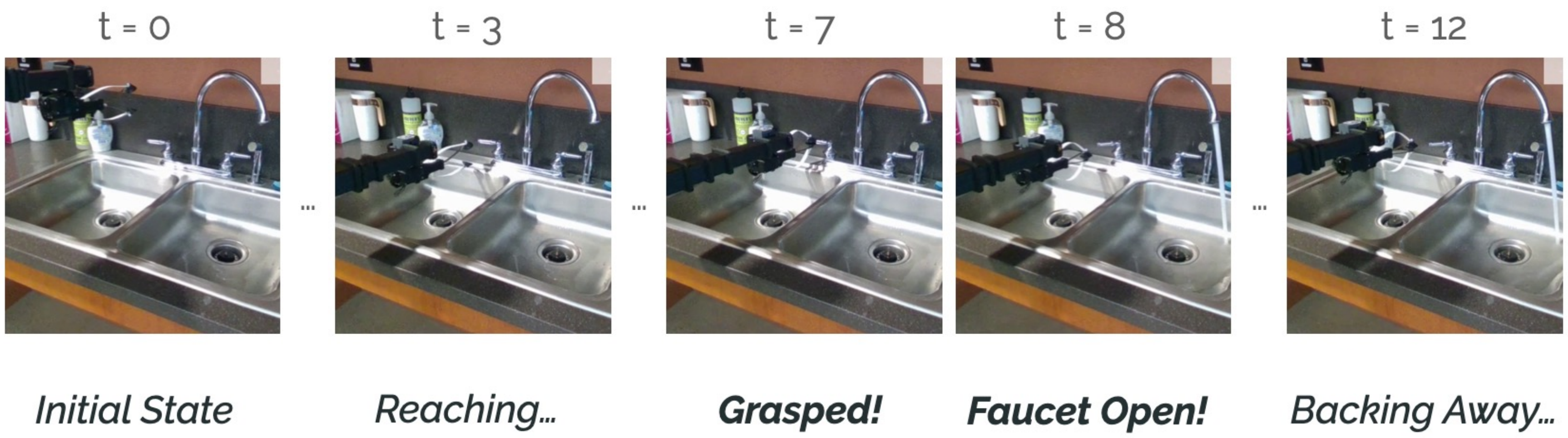


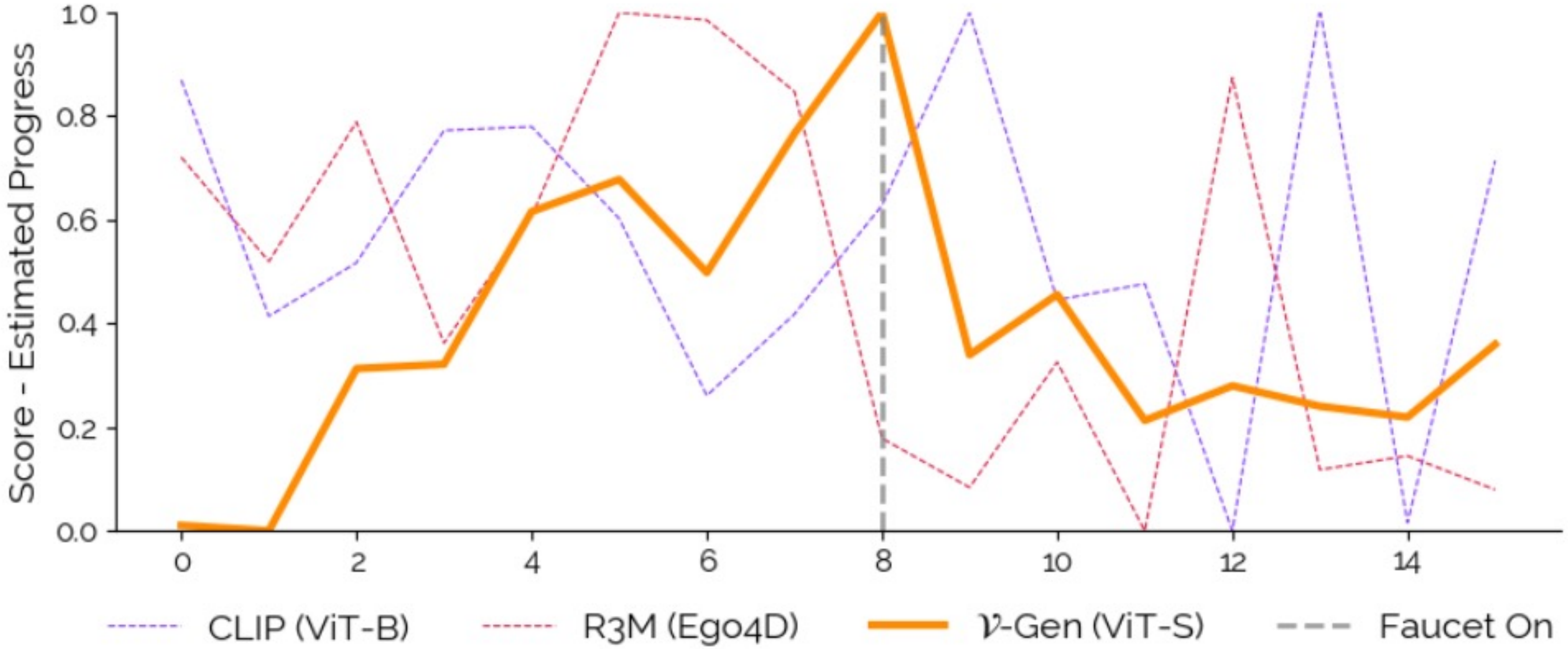
Training on 20 demonstrations

Qualitative Zero-shot Intent Scoring -- Human



Qualitative Zero-shot Intent Scoring -- Robot





Give it a Try!

Voltron

V-Evaluation

https://github.com/siddk/voltron-robotics

https://github.com/siddk/voltron-evaluation

`pip install voltron-robotics`

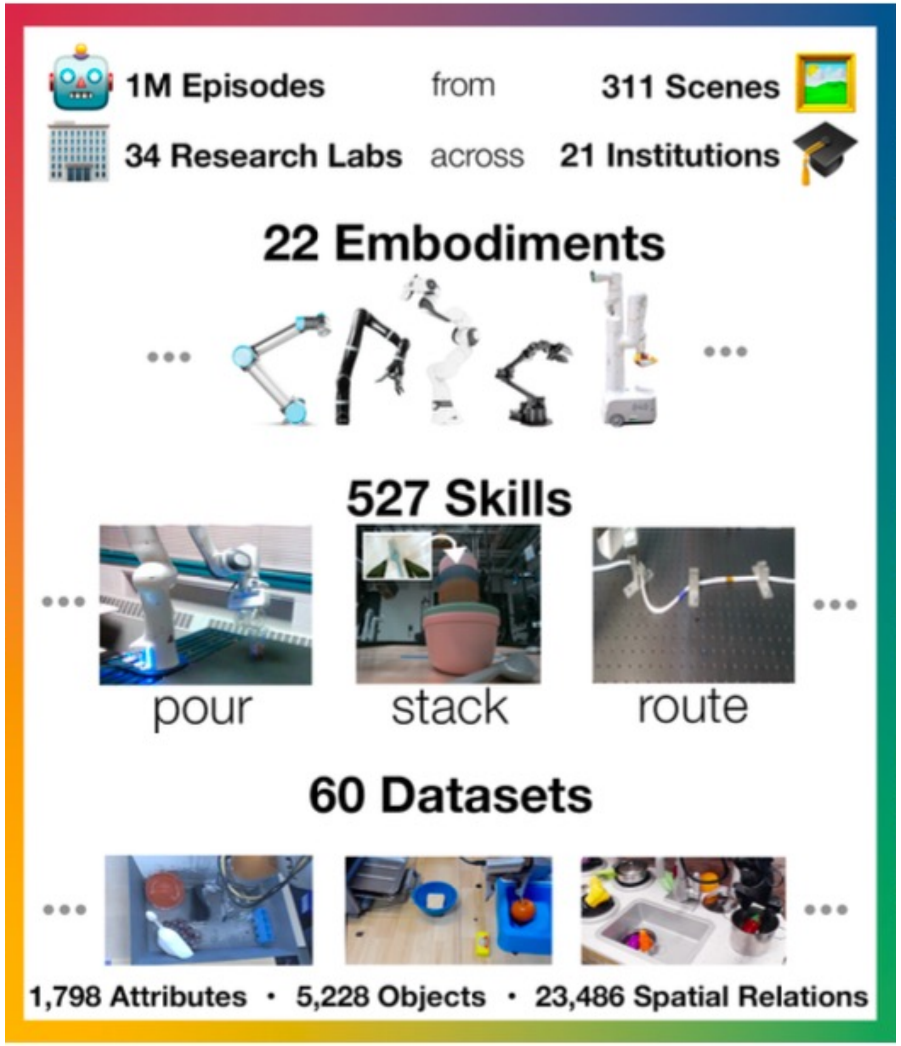
Language-Driven Representation Learning for Robotics
Siddharth, Karamcheti S. Nair, A. Chen, T. Kollar, C. Finn, D. Sadigh, P. Liang
arXiv preprint, February 2023

Key Takeaway 2

To tap into large offline datasets…

We should use language and multi-frame conditioning to integrate ***syntax***, ***semantics***, and ***pragmatics*** for learning visual representations useful for robotics.

Open X-Embodiment: Robotic Learning Datasets and RT-X Models



Take 1: What does it take to build a robotics foundation model?

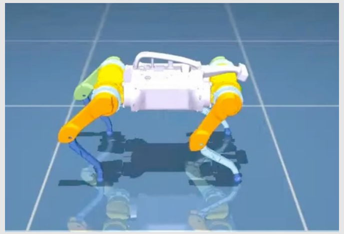
Instead of learning from preference queries or demonstrations, can we tap into large offline datasets?

Take 2: What are some ways of using existing pretrained large models?

Instead of learning from preference queries or demonstrations,
or tapping into large offline datasets.
can we tap into the existing knowledge of LLMs/VLMs?

Large Models enable …

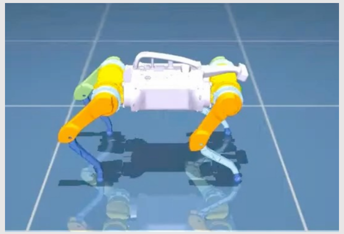
Reward Design



[Kwon et al, ICLR23] [Yu et al. CoRL23]

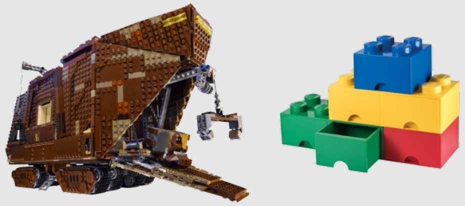
Large Models enable …

Reward Design



[Kwon et al, ICLR23] [Yu et al. CoRL23]

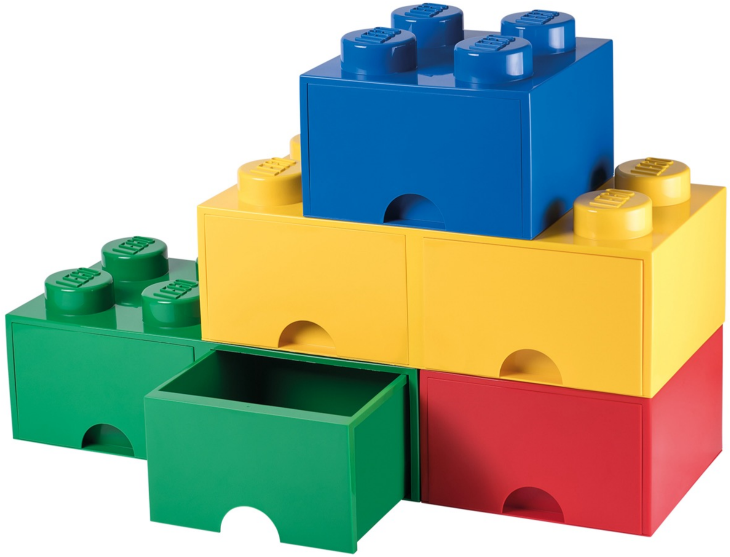
Commonsense Reasoning



[Kwon et al, in submission]

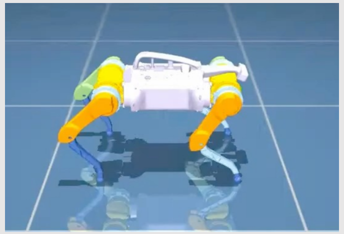

BRAVE &
STRONG

How to know not to clean
the **intricately built Legos** but to put away the **Mega Legos**?



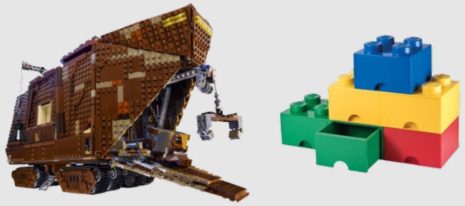
Large Models enable …

Reward Design



[Kwon et al, ICLR23] [Yu et al. CoRL23]

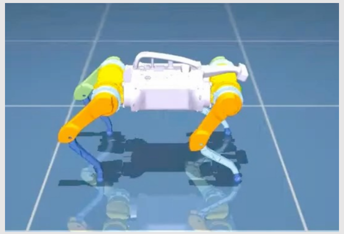
Commonsense Reasoning



[Kwon et al, in submission]

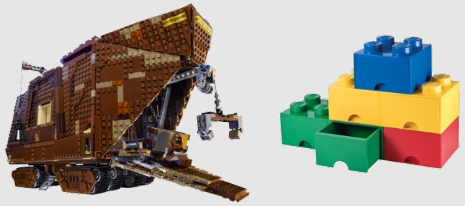
Large Models enable …

Reward Design



[Kwon et al, ICLR23] [Yu et al. CoRL23]

Commonsense Reasoning

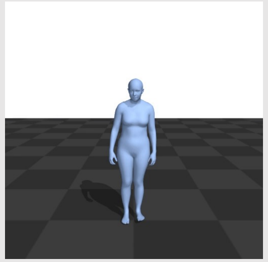


[Kwon et al, in submission]

Semantic Manipulation

[Sundaresan et al. CoRL23]

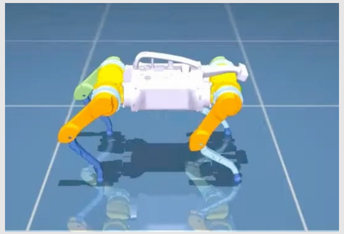
Teaching Humans



[Srivastava et al. ICML23]

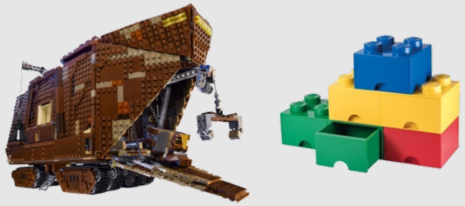
Large Models enable …

Reward Design



[Kwon et al, ICLR23] [Yu et al. CoRL23]

Commonsense Reasoning

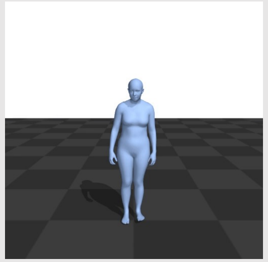


[Kwon et al, in submission]

Semantic Manipulation

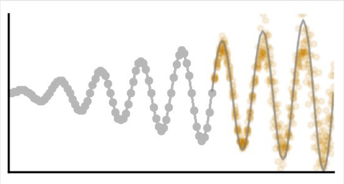
[Sundaresan et al. CoRL23]

Teaching Humans



[Srivastava et al. ICML23]

Pattern Machines



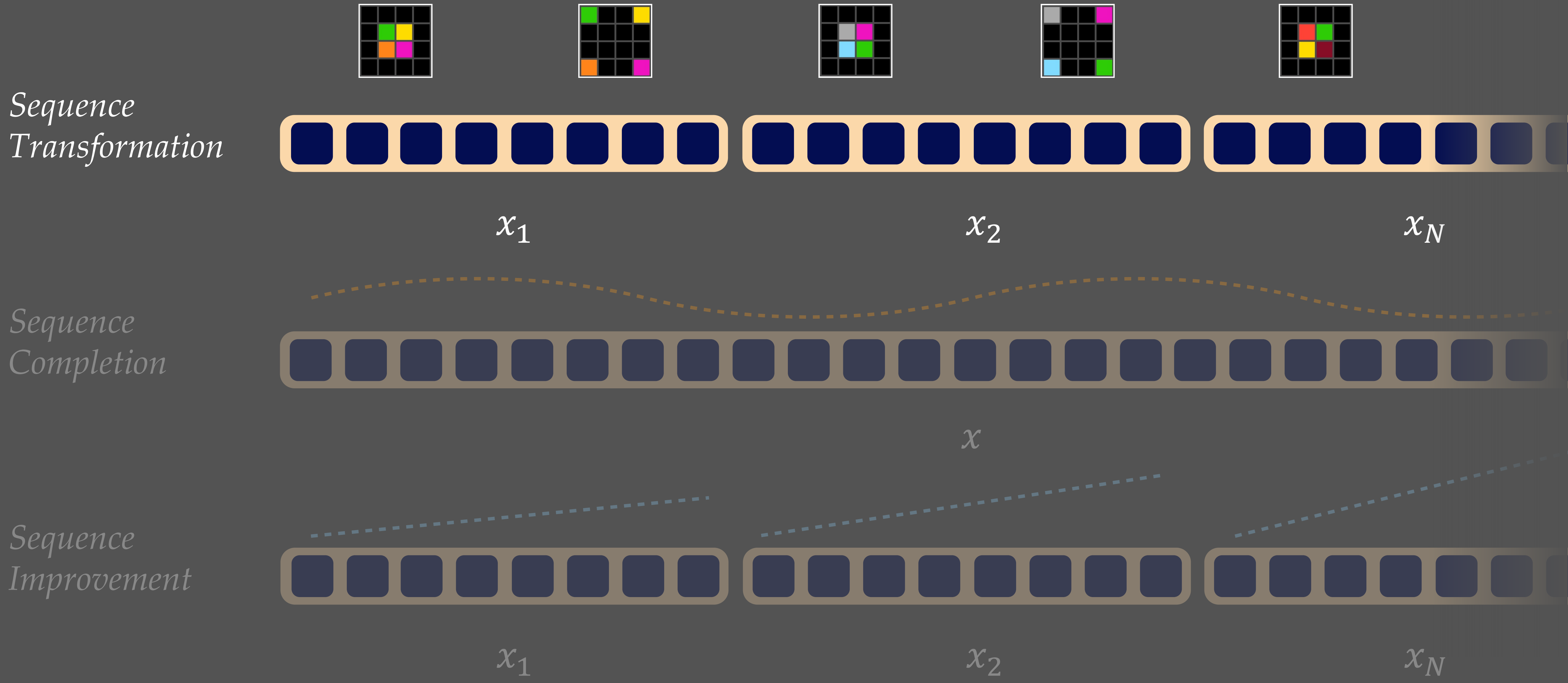
[Mirchandani et al. CoRL23]

We could go beyond leveraging LLMs understanding of semantics and context…

They're great pattern machines!

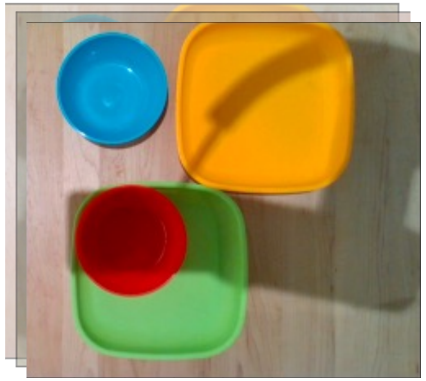
LLMs as General Pattern Machines (Mirchandani et al. 2023)

Suvir Mirchandani



Sequence Transformation

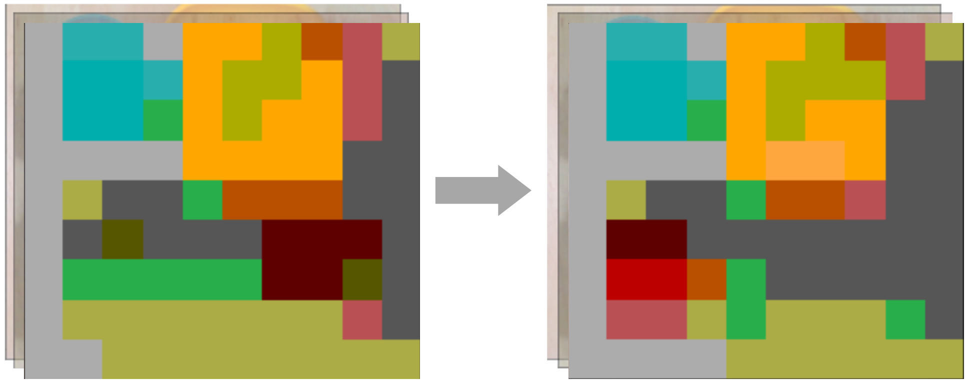
Train Examples



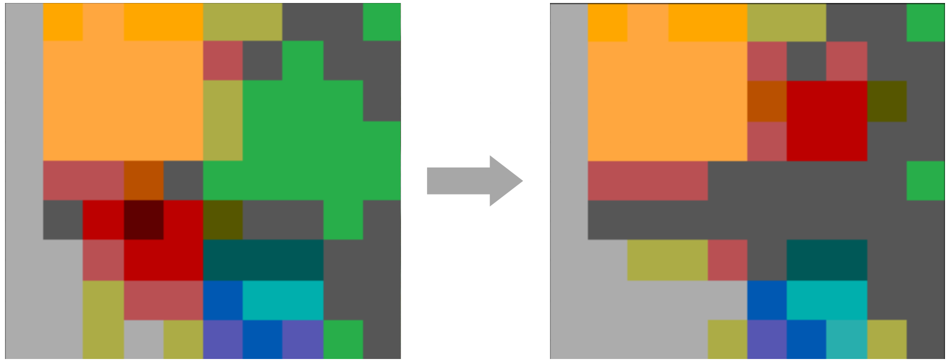
Test Example

Sequence Transformation

Train Examples



Test Example

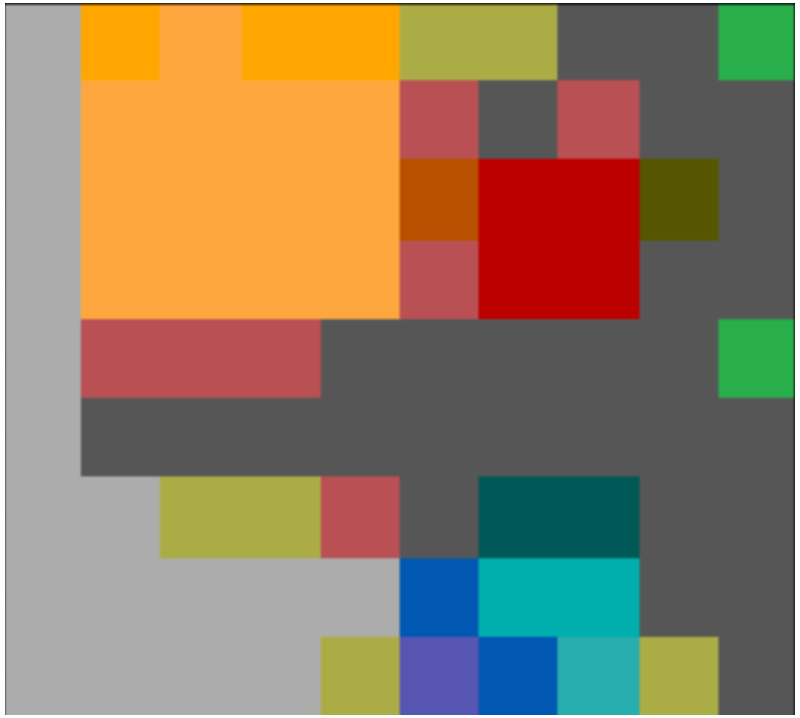


63 47 47 63 77 77

. . .

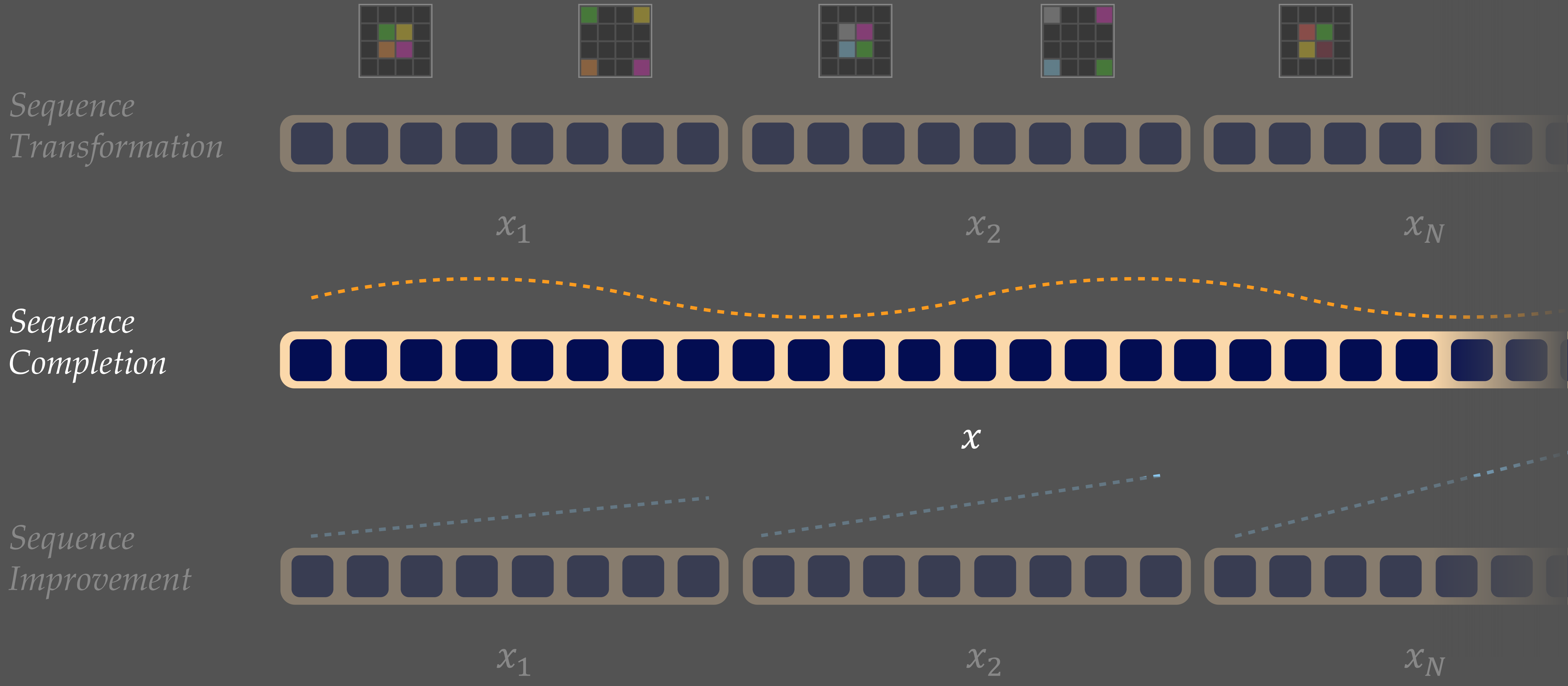
63 62 42 42 46 57
63 37 37 42 42 42
63 53 53 57 46 42
63 58 58 62 46 62

Sequence Transformation



LLMs as General Pattern Machines (Mirchandani et al. 2023)

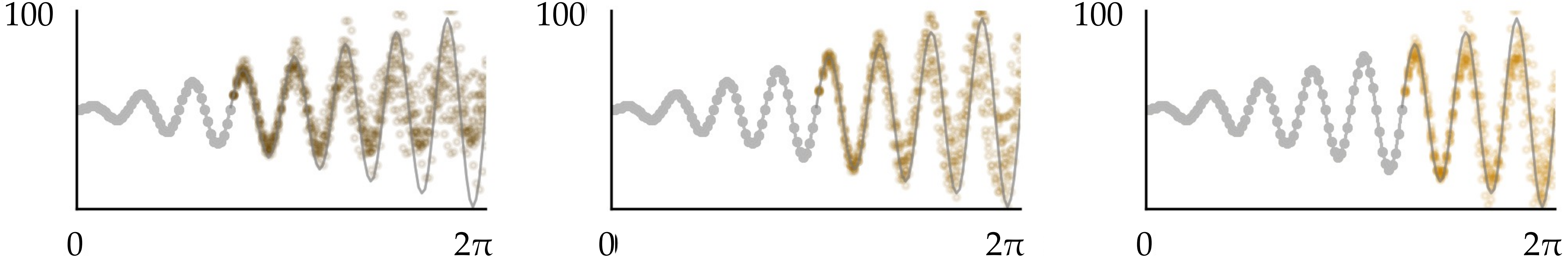
Suvir Mirchandani



Sequence Completion

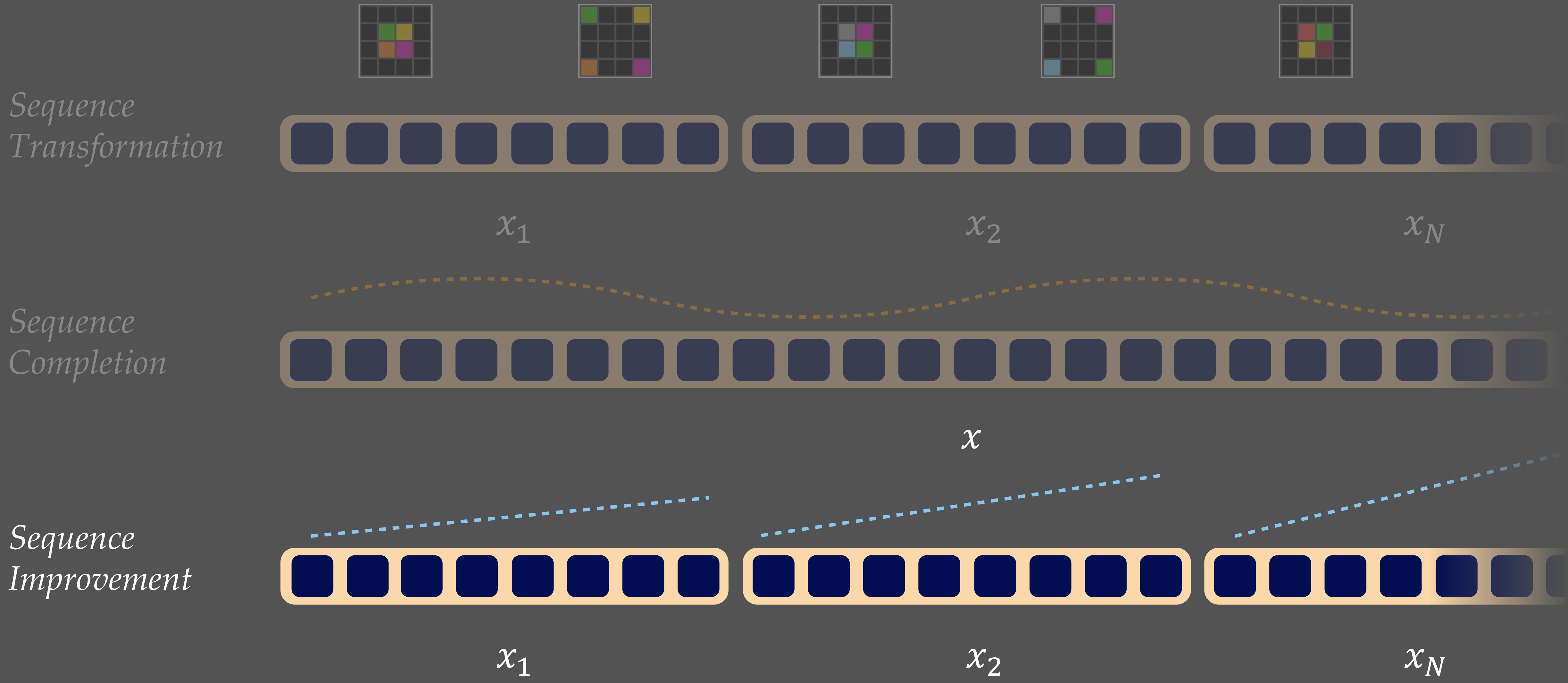
- Evaluate how well LLMs of various scales can extrapolate simple functions (e.g. sinusoids)

$$f(x) = ax\sin bx$$

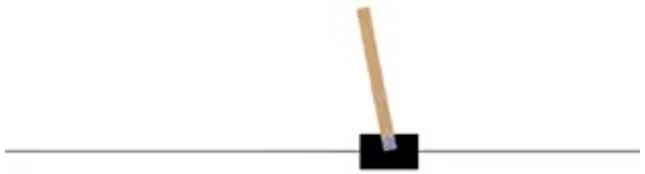


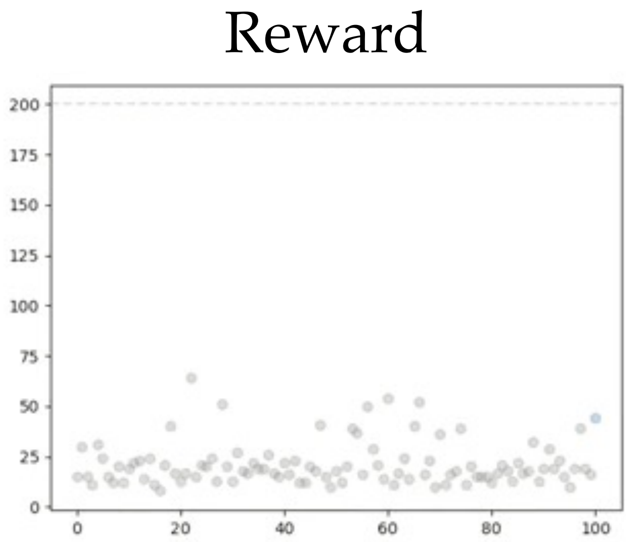
LLMs as General Pattern Machines (Mirchandani et al. 2023)

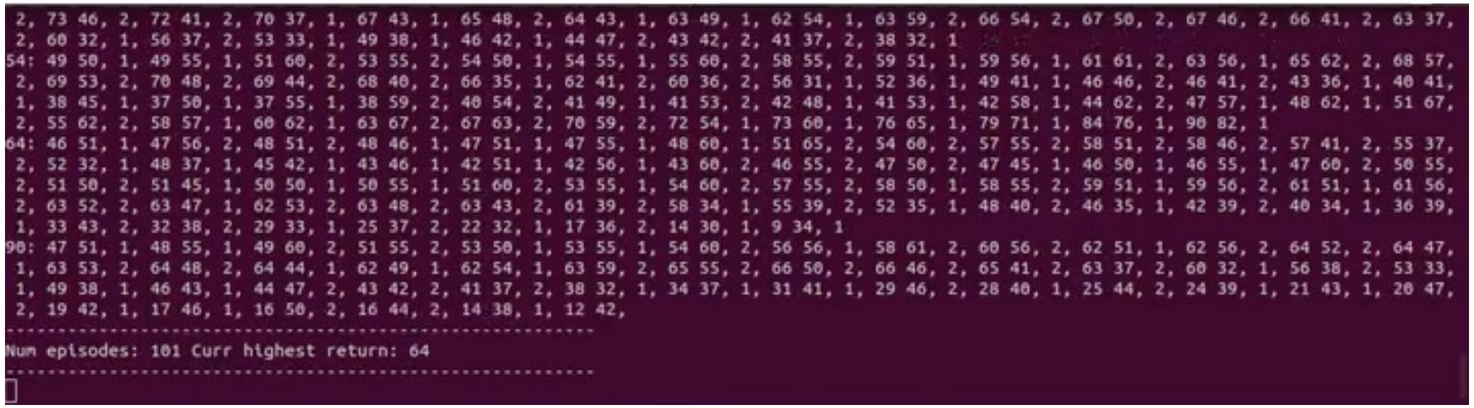
Suvir Mirchandani

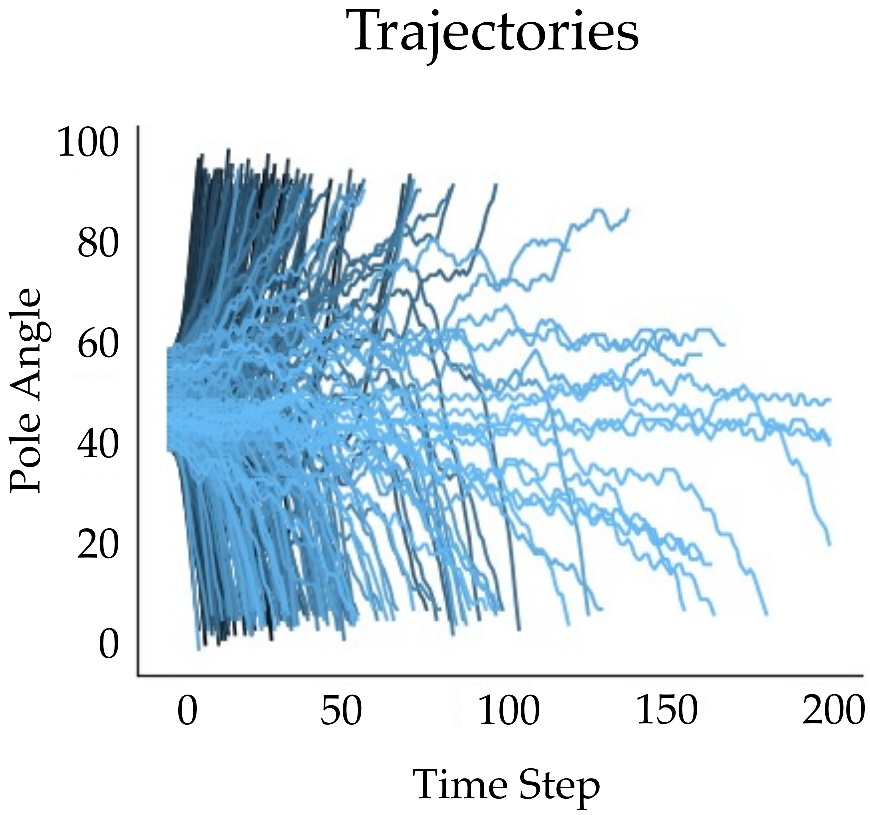


Sequence Improvement



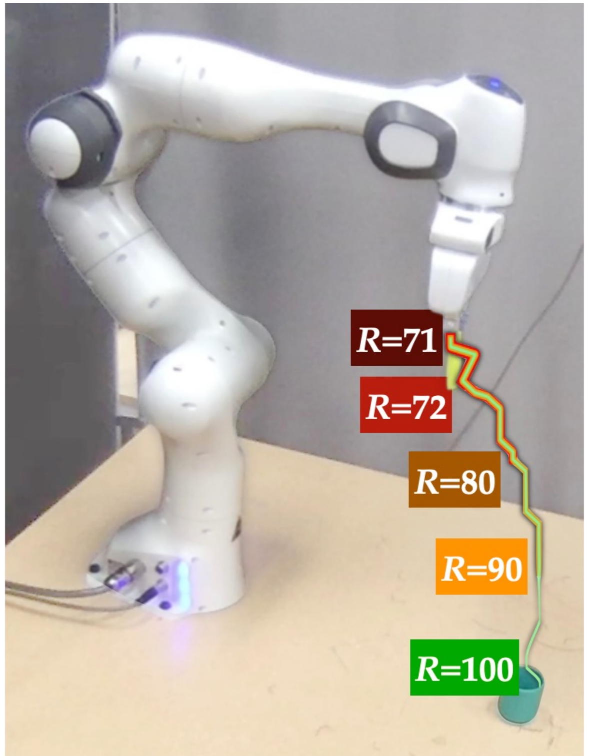






Sequence Improvement

We initialize the context with a series of trajectories, and prompt the LLM to produce a higher-reward trajectory



reward *trajectory*

```
71: 104 83 123, 104 83 123, ...
72: 104 83 123, 104 83 123, ...
80: 104 83 123, 104 83 123, ...
90: 104 83 123, 104 83 123, 104 83 123, 104 83 123, 104 83 123,
104 83 123, 104 83 123, 104 83 123, 104 83 123, 104 83 123, 104
83 123, 104 83 123, 104 83 123, 104 83 123, 104 83 123, 105 83
123, 105 83 123, 106 83 123, 106 83 123, 107 83 123, 108 83 122,
109 83 122, 110 83 122, ...
100: 104 83 123
```

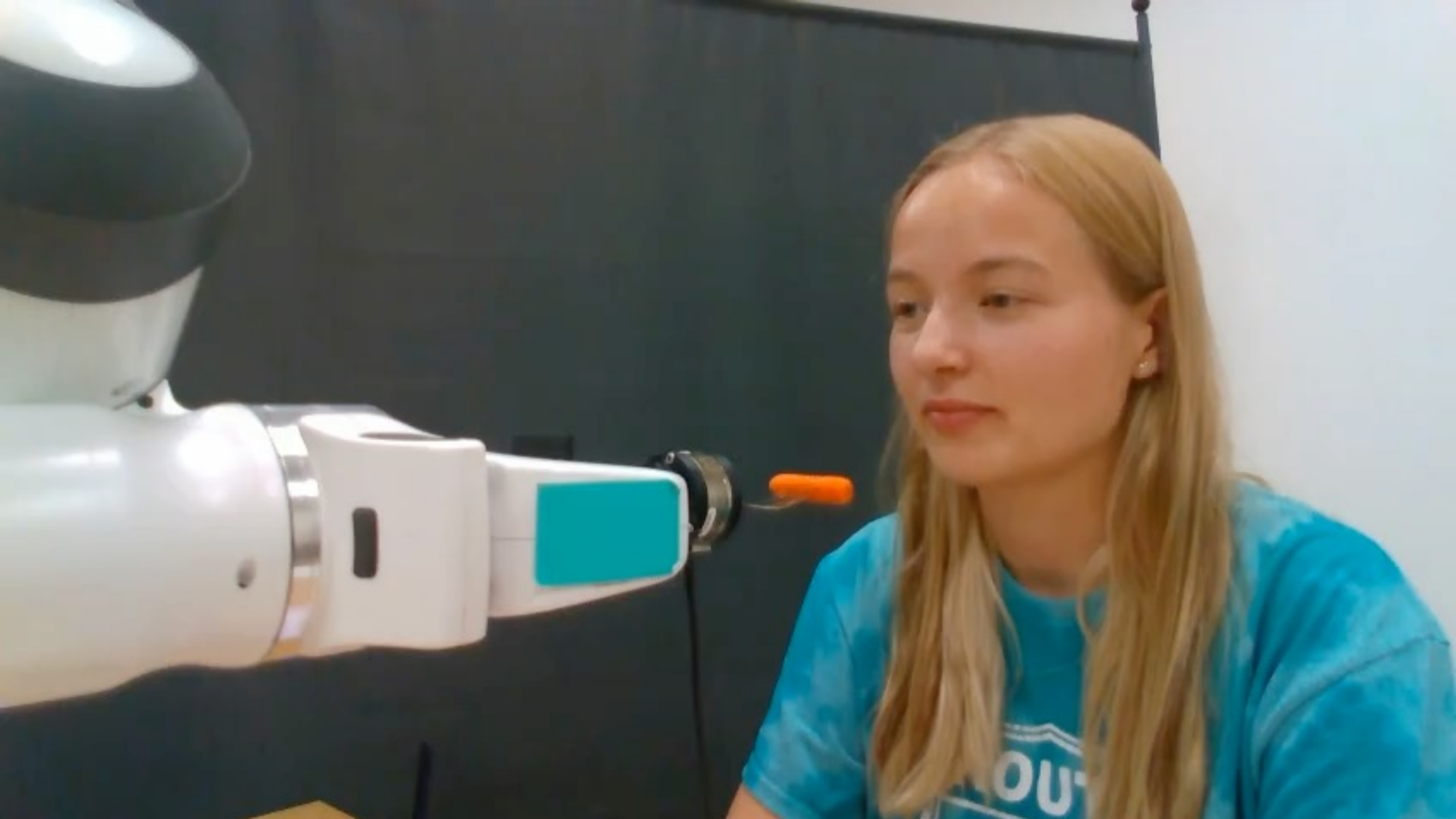
Clicker Training

Key Takeaway 3

LLMs not only enable reward design, social reasoning, semantic manipulation, and teaching humans

… but also can act as general pattern machines

… enabling sequence extrapolation, transformation, and optimization through the power of in-context learning.



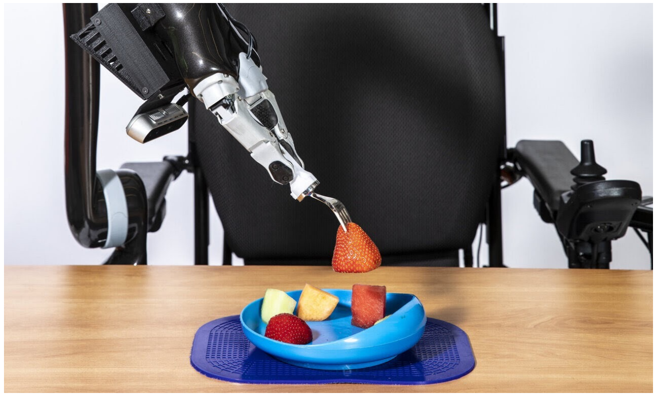
Robot-Assisted Feeding


Suneel Belkhale

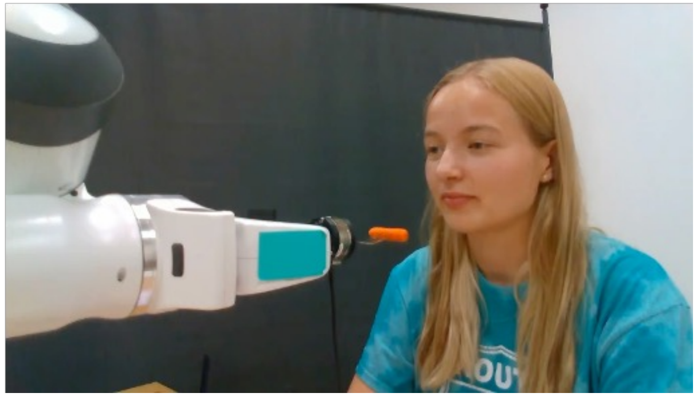

Priya Sundaresan


Jenn Grannen

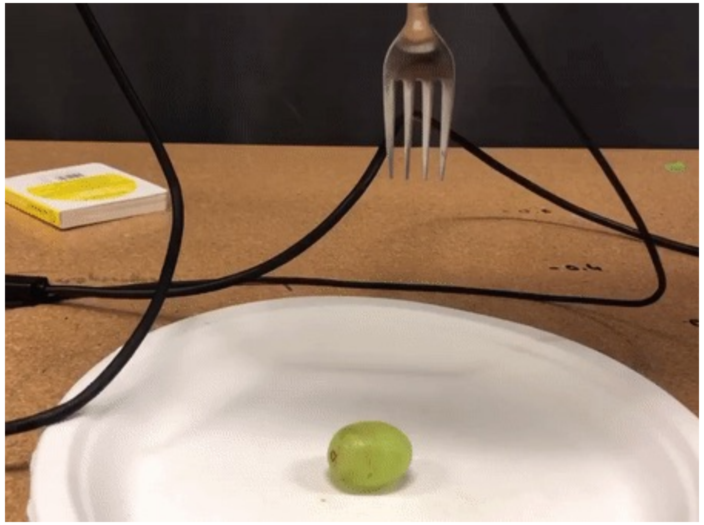
Acquisition:
Picking up food object

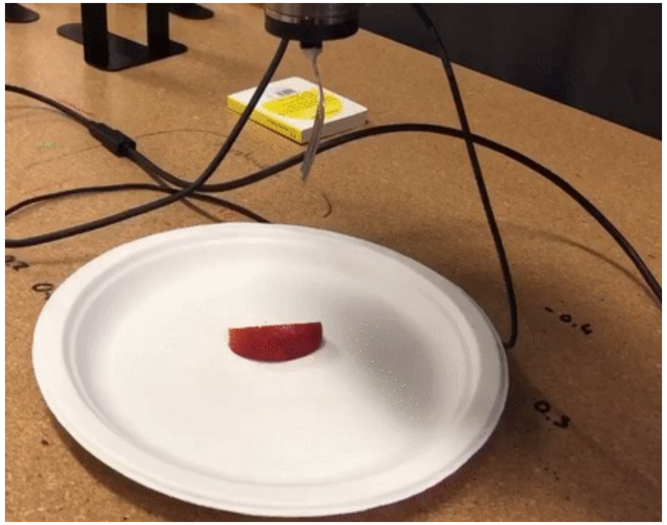


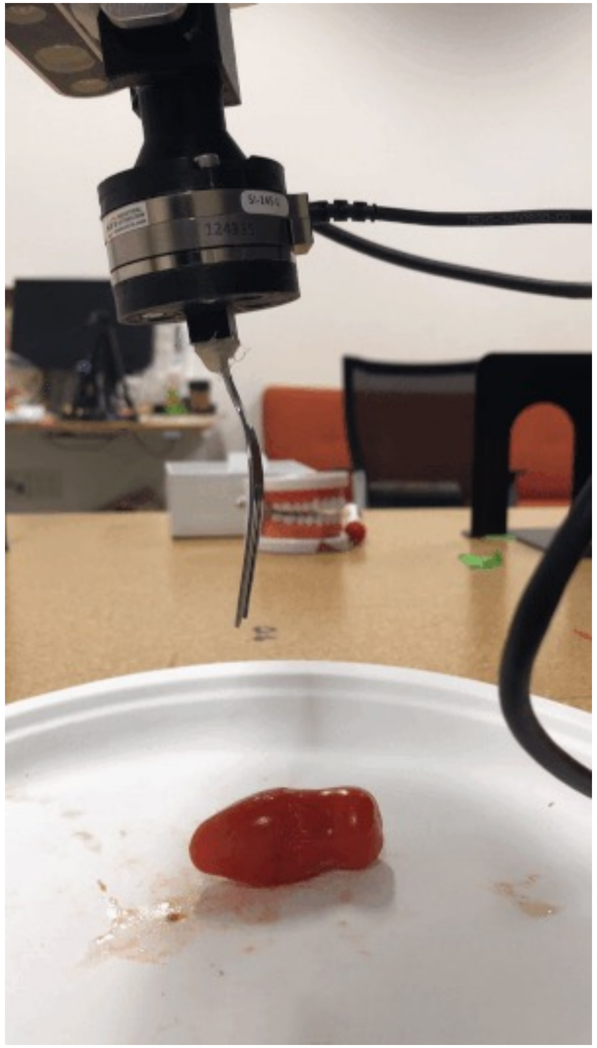
Bite Transfer:
Moving food into mouth

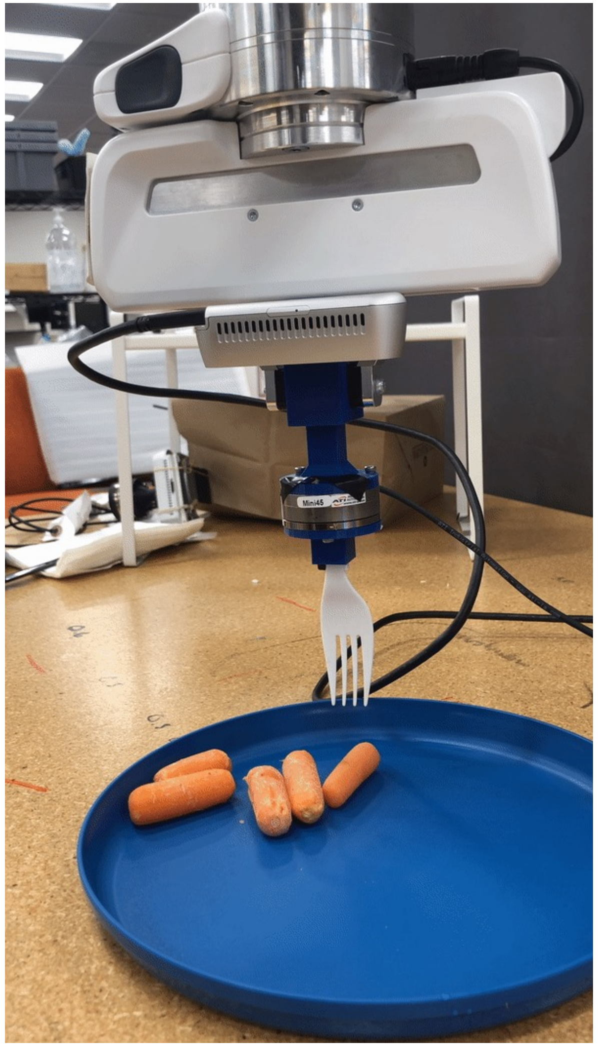


Bite Acquisition - Failures

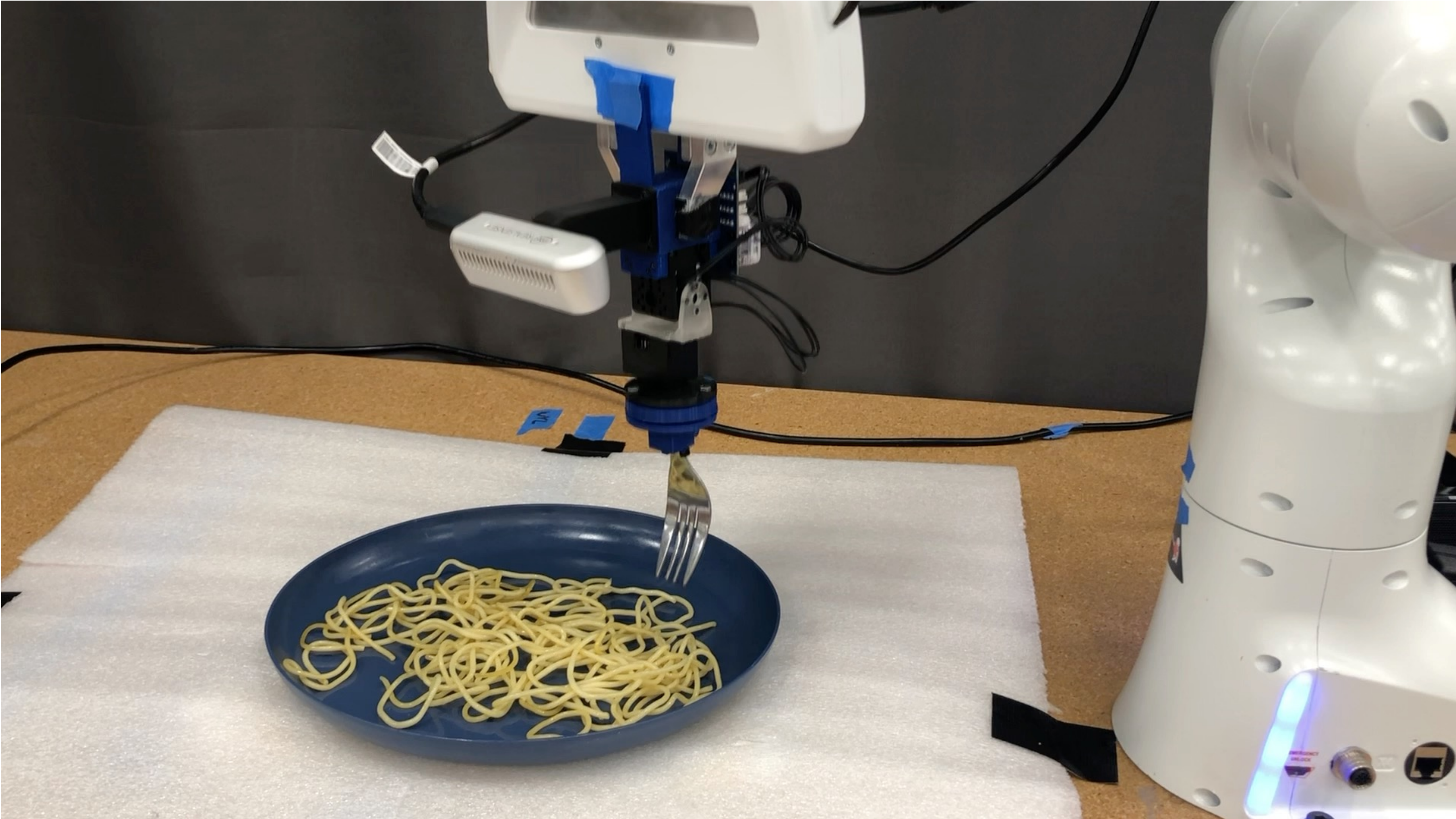


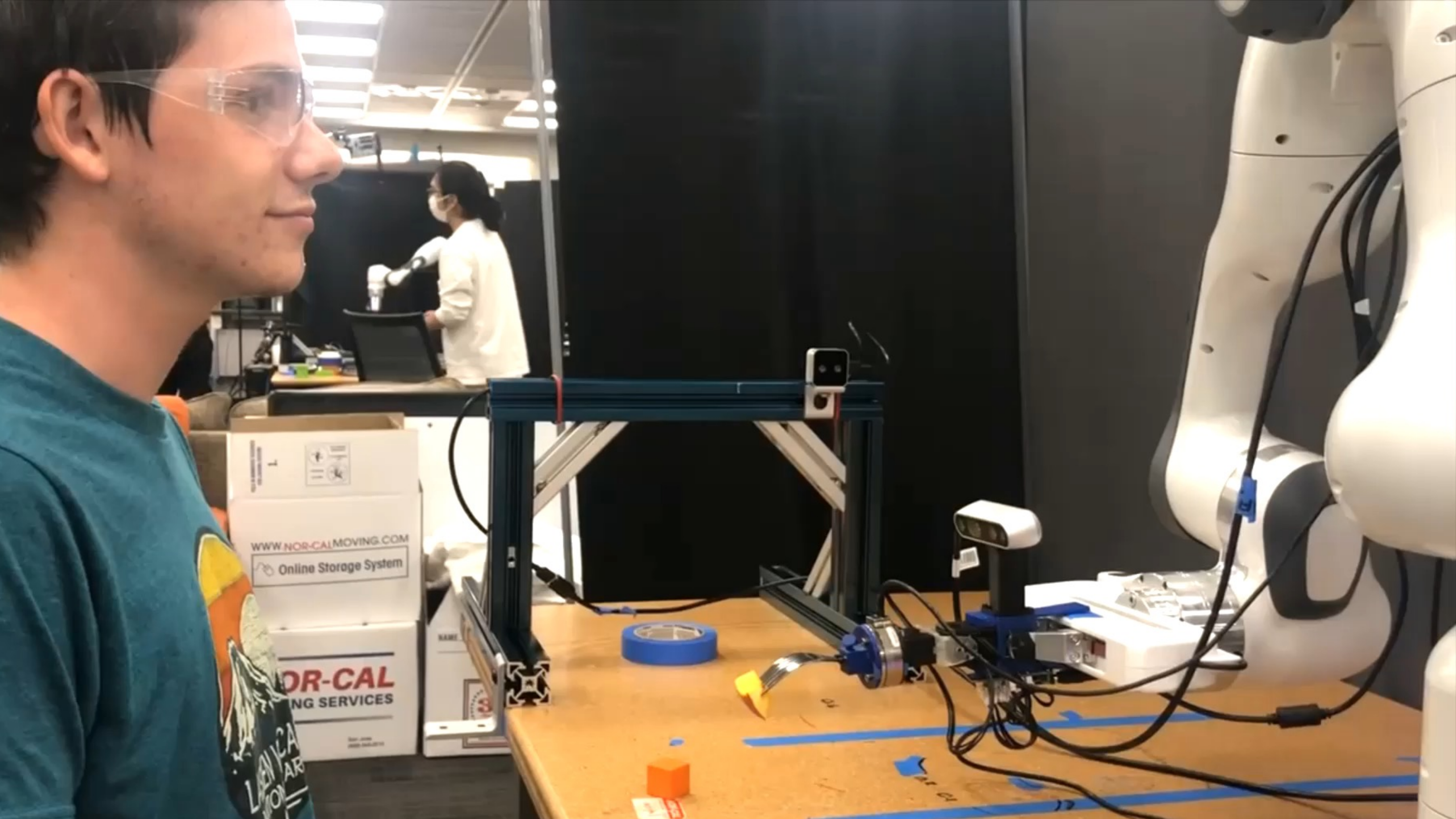




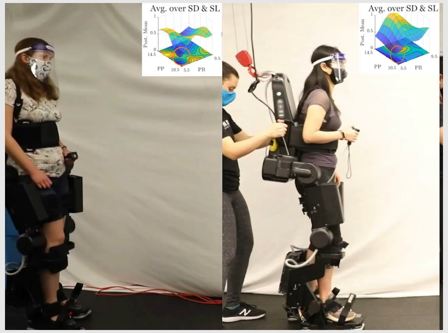


Leverage visual and haptic observations during interaction with an item to rapidly and reactively plan skewering motions




WWW.NOR-CALMOVING.COM
Online Storage System
OR-CAL
ING SERVICES

Learning Human Preferences



Biyik et al. IJRR 21
Kwon et al. ICLR 23
Gandhi et al. CoRL 22

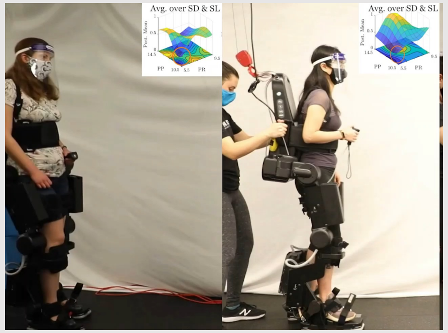
Foundation Models for Robotics

Voltron

Karamcheti et al. RSS23
Mirchandani et al. CoRL23


Stanford
ENGINEERING
STANFORD

Learning Human Preferences



Biyik et al. IJRR 21
Kwon et al. ICLR 23
Gandhi et al. CoRL 22

Foundation Models for Robotics

Voltron

Karamcheti et al. RSS23
Mirchandani et al. CoRL23