HRI with Ambiguous Intent

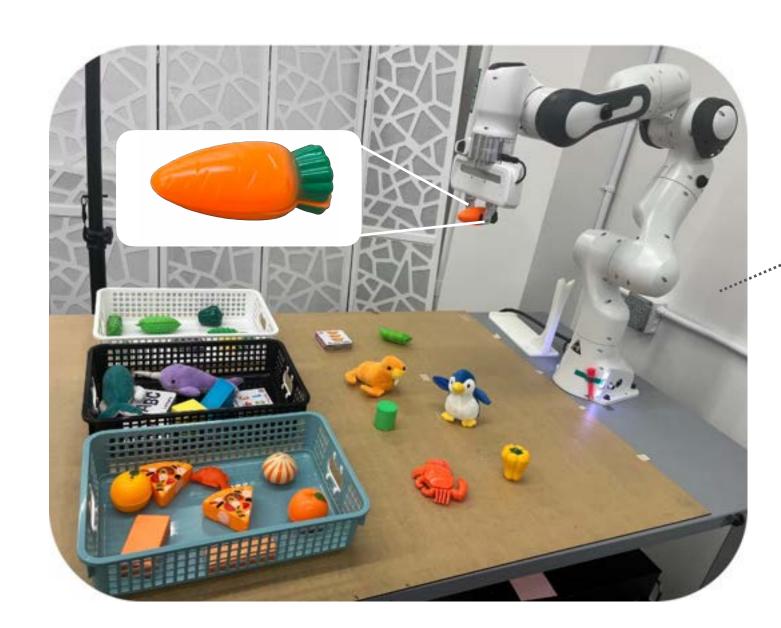
Certifiable Autonomy with RCIP

Contingency Planning

Human

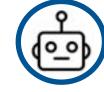


Clean up the table, please.



Discrete Set of Intents

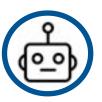
The human wants me to put the carrot with the...



[Vegetables] [Toys] [Orange Things]

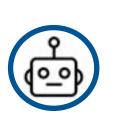
Prediction Model (e.g., VLM)

 $p_1 = 0.55$ $p_2 = 0.08$ $p_3 = 0.37$



Environment Context and State History

There are three bins of vegetables, toys, and orange: things. The human recently put more green produce with the vegetables. Now I have to sort the carrot.



Robot Observation(s)

I have an image of each bin and the carrot.

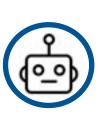


Predicted Action Set

0.55 - Put the carrot with the vegetables

0.37 - Put the carrot with the orange item

0.08 - Put the carrot with the toys



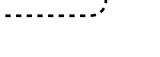
Temp. scaling

Prediction Set for Statistical Risk Calibration

Risk-Controlling Prediction Threshold: 0.38 Risk-Controlling Temperature Scaling: 0.05

0.51 - place(Vegetables)





Trigger Human Help to Reveal True Intent

Prediction set size 2 > 1

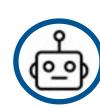




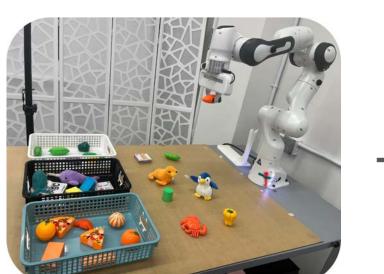
Put it with the [Orange Things]







Current Task State

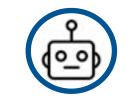








Intent-Conditioned Planner



place(Toys) place(Veg.) place(Orange)







Intent Satisfaction with State Constraint





