Introspective Planning: Aligning Robots' Uncertainty with Inherent Task Ambiguity

Kaiqu Liang, Zixu Zhang, Jaime F. Fisac

Princeton University

Background & Motivation

Ensure robot actions are both **safe** and **compliant**

Challenges?

- **Task ambiguity**
- **Guarantee safety**

What are the previous solutions?

Direct Prompting → Over-Step !

LLM hallucination

Conformal Prediction → Over-Ask !

Too conservative

How to refine robot's uncertainty?

Step 1: Construct the knowledge base — **Introspective reasoning generation**

Step 2: Retrieve most relevant knowledge — **Guide uncertainty + safety aware planning**

Apply conformal prediction after introspection + → **Tighter guarantee**

Quantitative Results

SR: Success Rate | HR: Help Rate | ESR: Exact Set Rate
OAR: OverAsk Rate | OSR: OverStep Rate | UR: Unsafe Rate
NCR: Noncompliance Contamination Rate | UCR: Unsafe Contamination Rate

Method	Prediction Metrics			Decision Metrics				
	ESR ↑	NCR ↓	UCR ↓	SR ↑	HR	OAR ↓	OSR ↓	UR ↓
KnowNo (Conformal)	37.5	51.0	7.0	84.5	77.5	51.3	35.3	1.0
Prompt Set	73.5	11.5	3.5	82.5	63.0	3.8	36.2	2.5
Prompt Set + CoT	79.0	10.0	5.0	87.5	67.0	10.3	30.8	4.0
Retrieval-Q-CoT	81.5	7.0	4.5	88.0	65.0	2.6	26.1	4.0
Auto-CoT	77.5	10.0	5.0	85.5	62.5	1.3	37.3	4.0
Ours (Conformal)	58.0	27.5	3.0	87.5	63.0	6.4	21.6	1.5
Ours (Direct)	**93.0**	**5.5**	**0.5**	**96.5**	67.5	**0.0**	**3.8**	**0.5**

Introspective planning generate more precise prediction sets

✓ avoids over-asking ✓ rarely oversteps ✓ safety-aware

Introspective planning + Conformal prediction

+ guarantee success + achieve tigher guarantee bound

Try introspective planning for your own planning task!

- **Uncertainty aware**
- **Guarantee safety**
- **Reflective reasoning**
- **Better interpretability**