Faithful Logical Reasoning via Symbolic Chain-of-Thought

Jundong Xu[1], Hao Fei[1], Liangming Pan[2], Qian Liu[3], Mong-Li Lee[1], Wynne Hsu[1]

[1] National University of Singapore, Singapore [2] University of California, Santa Barbara, USA [3] University of Auckland, New Zealand

TL;DR

We improve Chain-of-Thought prompting by combining symbolic representation and rules in logical reasoning.

Introduction

Question

Premises: There was one six-way tie on the leaderboard and one person in the six-way tie was from Belgium. Descampe is from Belgium and is on the leaderboard of the 1992 du Maurier Classic.

Is the following statement true, false, or uncertain? Descampe is in the six-way tie in the leaderboard of the 1992 du Maurier Classic.

Chain-Of-Thought

The context states that there was one six-way tie on the leaderboard and one person in the six-way tie was from Belgium. Descampe is from Belgium and is on the leaderboard of the 1992 du Maurier Classic. Therefore, it can be inferred that Descampe is in the six-way tie in the leaderboard of the 1992 du Maurier Classic. ✗ *overgeneralization*

Symbolic Chain-Of-Thought

Translated Symbolic Premises:

- ∃x (Tie(x, sixWay) ∧ From(x, Belgium)
- From(Descampe, Belgium) ∧ Leaderboard(Descampe, 1992DuMaurierClassic)

Reasoning Based on Symbolic Representation and Rules:

Since there is at least one person from Belgium in the six-way tie and Descampe is from Belgium, it is possible that Descampe is the person in the six-way tie. *Existential Instantiation ,Conjunction Elimination*

Evaluate the link using ***First-Order Logic*** from the step from ∃x (Tie(x, sixWay) ∧ From(x, Belgium)) to Tie(Descampe, sixWay) requires a premise directly stating Descampe's involvement in the six-way tie, which is absent such as Tie(Descampe, sixWay). However, such information is not given. Thus, the truth of the statement "Tie(Descampe, sixWay)" remains unknown. ✓

Figure 1: An illustrative example of logical reasoning via Chain-of-Thought and our proposed Symbolic CoT (SymbCoT).

Logical reasoning involves evaluating evidence, constructing arguments, and deducing concluons. While LLMs excel in understanding and generating human language, their performance in logical reasoning is limited by the complexities of logical structures and the need for precise, rule-based deductions.

Limitations of Current Techniques: Chain-of-Thought (CoT): Enhances LLM reasoning by mimicking human thought but struggles with rigorous logical rules. External Solvers (Logic-LM, LINC): Introduce errors and lose information during translation from natural language to symbolic representations.

Advantages of SymbCoT: **1)** Integrates symbolic reasoning within LLMs, combining symbolic expressions with natural language; **2)** Captures precise logical rules while retaining contextual nuances; **3)** Eliminates the need for external systems, reducing information errors and loss.

SymbCoT Framework

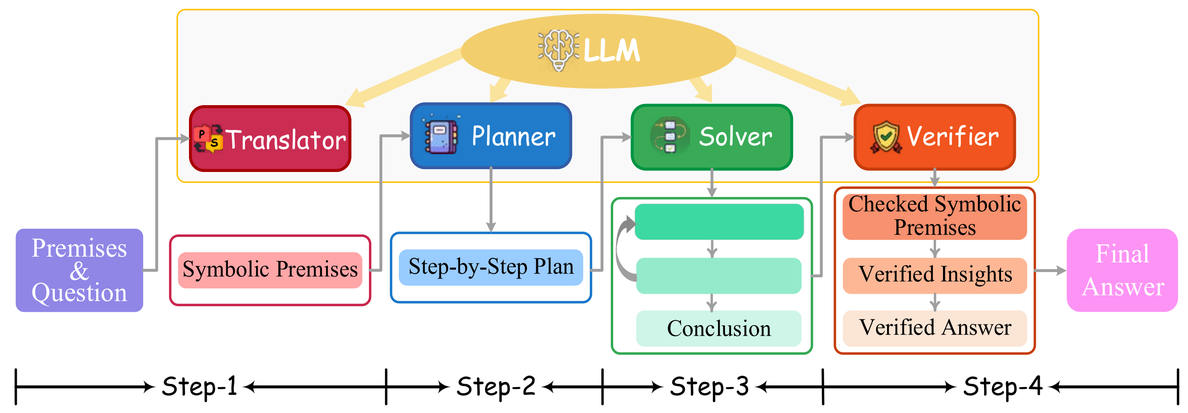

Figure 2: Overview of the workflow in our proposed symbolic CoT framework.

SymbCoT has four main modules: Translator, Planner, Solver, Verifier

Step 1. **Translator** translats natural language context into symbolic

```
Please parse the context P and statement S into
First-Order Logic formulas
```

Step 2. **Planner** derives plan based on the natural and symbolic context (P_c and S_c)

```
Please derive a step-by-step plan using the First-Order
Logic rule for determining the conclusion based on the
context P_c and S_c
```

Step 3. **Solver** solves the problem given context and plan

```
Given P_c, S_c and I, Please solve the question based on
First-Order Logic rules
```

Step 4. **Verifier** verifies the translation and solving process

```
Given P_c, S_c D and C, Please verify:  1) the symbolic
context P' and S' is consistent with the natural
language P and S; 2) the solving step D is logically
valid
```

Experiment Results and Analysis

	ProntoQA	ProofWriter	FOLIO	Avg
• *GPT-3.5-turbo*				
Naive	47.40	35.50	45.09	42.66
CoT	67.80	49.17	57.35	58.11
Logic-LM	61.00	58.33	**62.74**	60.69
SymbCoT	**75.80**	**59.03**	57.84	**64.22**
	(+8.00)	(+0.70)	(-4.90)	(+3.53)
• *GPT-4*				
Naive	77.40	52.67	69.11	66.39
CoT	98.79	68.11	70.58	79.16
CoT-SC	-	69.33	68.14	-
ToT	-	70.33	69.12	-
CR	-	71.67	69.11	-
DetermLR	-	79.17	75.45	-
Logic-LM	83.20	79.66	78.92	80.59
SymbCoT	**99.60**	**82.50**	**83.33**	**88.47**
	(+0.81)	(+2.84)	(+4.41)	(+7.88)

Table 1: Performance on symbolic reasoning with First-Order Logical representation.

	LogicalDeduction	AR-LSAT	Avg
Naive	71.33	33.33	52.33
CoT	75.25	35.06	55.14
CoT-SC	74.67	-	-
ToT	76.83	-	-
CR	78.33	-	-
DetermLR	85.00	-	-
Logic-LM	87.63	43.04	65.34
SymbCoT	**93.00**	**43.91**	**68.46**
	(+5.37)	(+0.87)	(+3.12)

Table 2: Results (using GPT-4) on symbolic reasoning with Constraint Optimization representation.

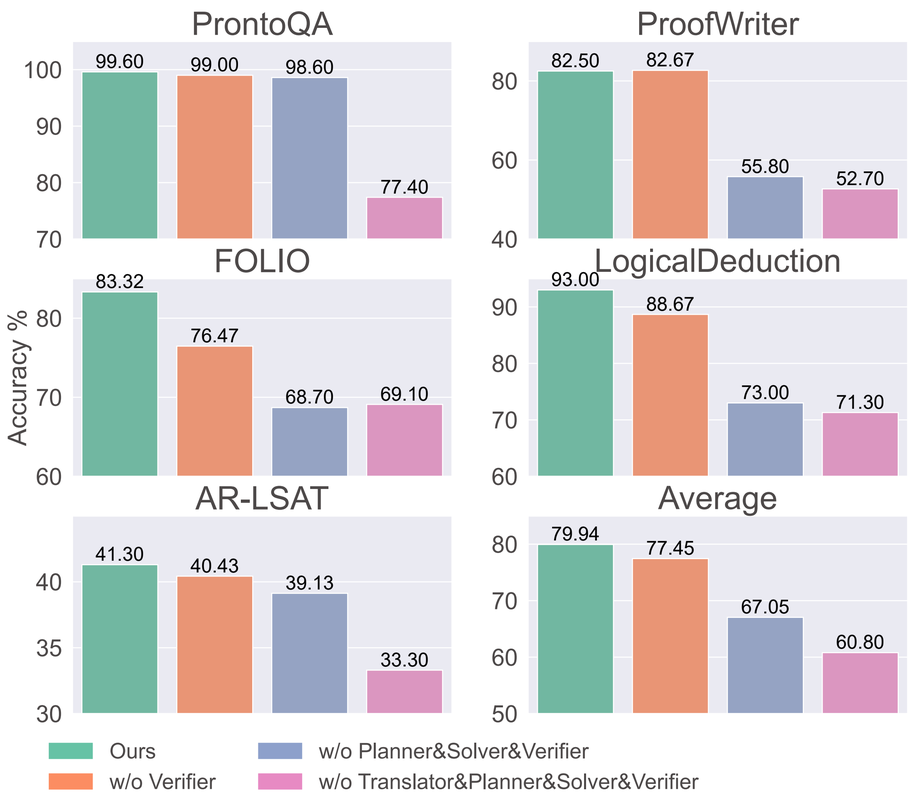

Figure 3: Ablation study. Since the Solver is dependent on the Planner, they have to be ablated simultaneously.

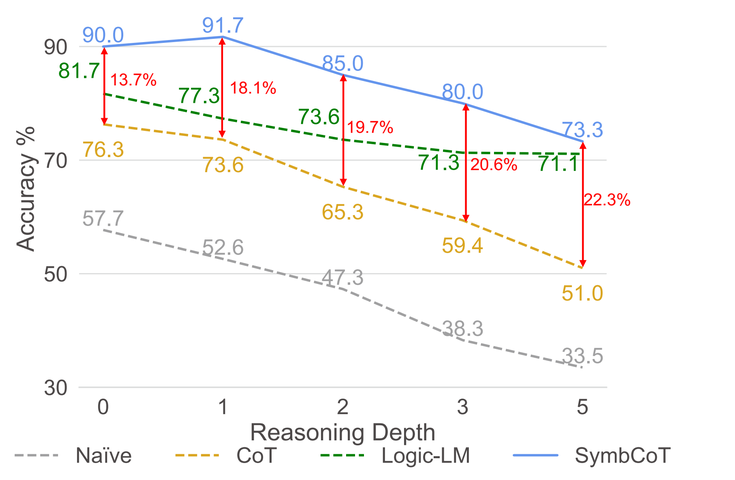

Figure 4: The effect of reasoning depth with GPT-4 on ProofWriter. The red dual-head arrow indicates our improvements over vanilla CoT.

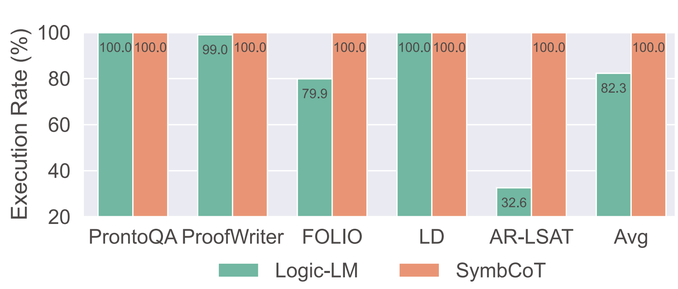

Figure 5: Execution rate between Logic-LM and Ours.

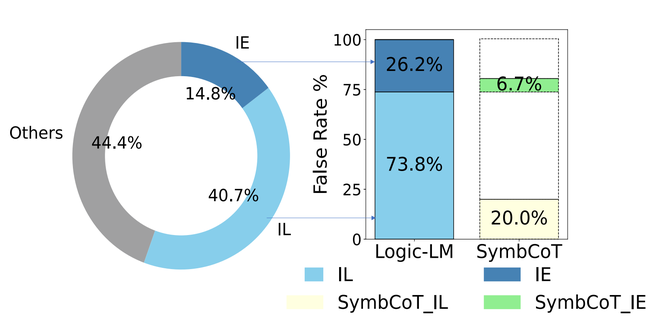

Figure 6: The left pie shows the error proportion from the external solver due to 1) Information Loss (IL), 2) Information Error (IE), and Others. The bar chart consists of two parts. The left bar shows the false rate from the external solver made by IL/IE adding up to 100%. The right bar shows the reduced false rates via our method.

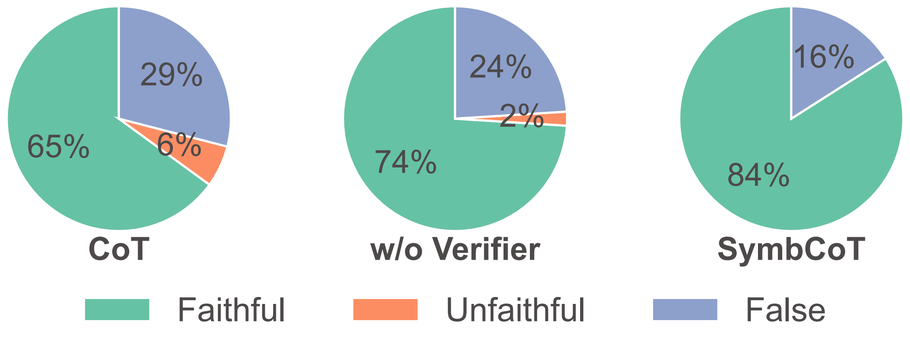

Figure 7: The proportion of faithful, unfaithful, and false answers. Faithful/unfaithful denotes whether the predicated correct answer is derived from valid&reaonsable logical reasoning.

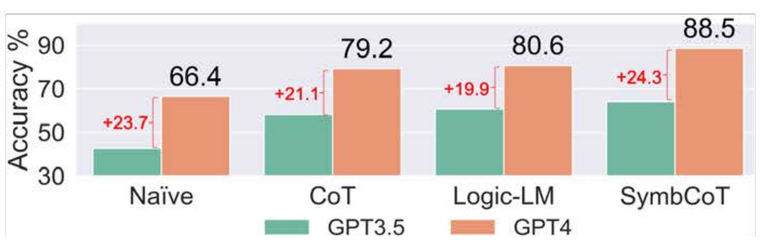

Figure 8: The Improvement from GPT-3.5 to GPT-4.

Code

Paper