Empowering LLMs with Symbolic Representation and Reasoning

Fengxiang Cheng

Institute of Logic, Language and Computation, University of Amsterdam
f.cheng@uva.nl

Abstract

Large language models (LLMs) have achieved remarkable success in natural language processing tasks but still struggle with complex causal and logical reasoning. Previous neuro-symbolic methods can be summarized into a two-stage framework: first translating natural language (NL) problems into symbolic language (SL) representation, and then performing the symbolic reasoning process. To facilitate this direction, we provide a comprehensive survey, summarizing two main challenges including complex logical question-answering (QA) and cross-question logical consistency, and further propose a new taxonomy. To achieve precise symbolic representation and enhance the accuracy of LLMs' logical reasoning, we propose several effective and efficient approaches, including adaptively selecting the most suitable SL for each QA problem, a data-driven approach to determine the fine-tuning samples order, and an efficient multi-agent debate framework with sparse communication. Our future research will focus on theoretical analysis for optimal SL selection, translation refinement and robust neuro-symbolic approach to improve LLMs' reasoning.

Introduction

Large language models (LLMs) have achieved impressive success in various natural language tasks, but still face significant challenges in the complex causal and logical reasoning abilities, limiting their applicability in real-world scenarios. Previous methods to improve logical reasoning abilities of LLMs generally follow a two-stage framework: translating natural language (NL) problems into symbolic language (SL) and then performing reasoning using these symbolic representations (Cheng et al. 2025b). The first **NL-to-SL translation stage** leverages LLM prompting, while the subsequent **logical reasoning stage** can be executed via external solvers (Ye et al. 2023), further LLM prompting (Xu et al. 2024), or fine-tuned models (Morishita et al. 2024). However, current methods are often limited in two aspects. On the one hand, translation results are prone to errors or information loss, which significantly impacts subsequent reasoning. On the other hand, how to efficiently enhance the accuracy of LLMs' reasoning performance remains a significant challenge. To address these issues, my PhD research aims to the following research questions (RQs):

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

RQ1: How to comprehensively summarize and categorize the recent achievements and remaining challenges in LLM logical reasoning? We provided a comprehensive survey of state-of-the-art methods to improve LLMs' logical reasoning abilities including accurately solving each logical QA task and ensuring logical consistency across outputs.

RQ2: How to achieve precise symbolic representation in NL-to-SL translation stage to support the subsequent reasoning? We propose a method to adaptively select the most suitable SL for each logical QA problem before translation, achieving remarkable improvements in logical QA performance of LLMs. Our future work will focus on the theoretical analysis of optimal SL selection, as well as on how to further verify and refine the translation results of LLMs.

RQ3: How to develop effective and efficient algorithms to perform logical reasoning using the symbolic representation? We propose a data-driven approach to determine the fine-tuning sample order to enhance LLMs' reasoning abilities. In addition, we introduce a multi-agent debate framework with an adaptive sparse communication strategy to ensure both the effectiveness and efficiency of enhancing LLMs' logical reasoning abilities. Our future work will explore how to leverage the complementary advantages of different reasoning methods to further improve reasoning performance.

In a nutshell, my previous research has centered on causal and logical reasoning in language models. For future research, my key objective is to further improve performance of LLMs in both symbolic representation and reasoning.

Previous Work and Contributions

The translation and reasoning process generated by LLMs usually relies on their training corpus, leading LLMs to follow data-based correlation rather than rule-based inference. Since explicit data for complex causal and logical reasoning remains limited, language models frequently suffered from hallucinations stemming from spurious correlations (Cheng 2025). To mitigate this problem, we employ the probabilities of necessity and sufficiency for identifying causal and non-causal words in sentiment classification tasks. Then, we train a robust sentiment classification model using only the causal words while excluding the non-causal words in each sentence (Cheng et al. 2025a). Moreover, we further introduce a counterfactual contrastive learning approach to achieve more outperforming performance (Cheng et al. 2025c).

In response to RQ1, to our knowledge, we firstly provide a comprehensive survey of the logical reasoning of LLMs (Cheng et al. 2025b). We summarize and categorize two primary challenges: (1) accurate logical QA—it's difficult for LLMs to correctly answer complex logical reasoning problems which require complicated deductive, inductive, or abductive reasoning with given information, and (2) logically consistent outputs—LLMs often generate responses contradicting themselves *across* different questions. To facilitate this research direction, we investigate the cutting-edge approaches and introduce a detailed taxonomies for classifying them. Specifically, methods to improve LLMs' logical QA performance are categorized into solver-based, prompt-based, and fine-tuning. For logical consistency, we distinguish several categories involving negation, implication, transitivity, factuality and compositional consistency, as well as their corresponding evaluation metrics and solutions.

In response to RQ2, for the first time we reveal that the translation results and solutions of the same logical reasoning problem can vary significantly when expressed and solved in different SLs (Wang et al. 2026). Thus we propose a method to adaptively select the most suitable SL for each problem in the dataset before the NL-to-SL translation stage. Specifically, for each problem, we prompt LLMs to choose the optimal SL from first-order logic (FOL), logic programming (LP), and Boolean satisfiability (SAT) formulations based on their expressive features. Then, we leverage LLMs to translate the problem into the chosen SL and employ the corresponding solver to derive the final answer, showing that our adaptive-SL selection method outperforms other single-SL baselines.

In response to RQ3, we propose approaches to enhance LLMs' logical reasoning abilities while ensuring efficiency. By noting that LLMs' performance varies with the order of training samples during fine-tuning, we automatically determine samples' fune-tuning order according to their logical reasoning complexity (Cheng et al. 2026). Subsequently, we stratify the training dataset of propositional logic into several subsets and then we perform phased fine-tuning on them from low to high reasoning complexity, demonstrating remarkable performance on logical reasoning benchmarks and generalization on other reasoning tasks.

Moreover, we introduce a multi-agent debate approach with an adaptive sparse communication strategy to improve LLMs' logical reasoning abilities and ensure efficiency (Fu et al. 2025). Specifically, we prune unnecessary interactions based on confidence and information gains and allow agent update their memories with others' most valuable outputs, which reduces token costs by 25% in our experiments.

Directions for Future Work

For RQ2, our future work will concentrate on the theoretical analysis of optimal SL selection by logically formalizing the premises and statement, examining the expressiveness of different SLs for translation, as well as their computability for reasoning. In addition, we will explore effective approaches for translation verification and refinement by developing automated error-correction method with logical solvers that enforce both semantic and syntactic consistency.

For RQ3, our future work will explore neuro-symbolic approaches that leverage the complementary advantages of both solver-based reasoning and LLM-based reasoning to enhance logical reasoning performance while robust to translation errors. These two reasoning methods can benefit each other that logical solvers provide verifiable symbolic reasoning but are sensitive to translation errors, while LLM-based reasoning may result in hallucination but are more robust to imperfect translation. Our next step is to develop a method that leverages the complementary strengths of both approaches.

Conclusion

Overall, our work began by comprehensively surveying the main challenges of LLMs' logical reasoning abilities including logical QA and logical consistency. To achieve precise NL-to-SL translation and enhance the accuracy of LLMs' logical reasoning process, we propose several effective approaches, including adaptive SL selection for each problem, phased fine-tuning based on a reasoning complexity approach, and the efficient multi-agent debate framework with sparse communication. Our future research will focus on further theoretical analysis for optimal SL and developing robust neuro-symbolic approach to improve LLMs performance in both symbolic representation and reasoning.

References

Cheng, F. 2025. Enhancing the Logical Reasoning Abilities of Large Language Models. In *IJCAI*.

Cheng, F.; Li, H.; Leidinger, A.; and Van Rooij, R. 2025a. Revealing the Limitations of Exploiting Causal Effects to Resolve Linguistic Spurious Correlations. In *AAAI Workshop on AICT*.

Cheng, F.; Li, H.; Liu, F.; van Rooij, R.; Zhang, K.; and Lin, Z. 2025b. Empowering LLMs with logical reasoning: A comprehensive survey. In *IJCAI*.

Cheng, F.; Zhou, C.; Li, X.; Leidinger, A.; Li, H.; Gong, M.; Liu, F.; and van Rooij, R. 2025c. Mitigating Spurious Correlations via Counterfactual Contrastive Learning. In *Findings of EMNLP*.

Cheng, F.; Zhou, C.; Liu, F.; and van Rooij, R. 2026. Fine-Tuning Sample Order Matters in Propositional Logical Question-Answering (Student Abstract). In *AAAI*.

Fu, W.; Yang, H.; Cheng, F.; and Liu, F. 2025. Efficient First-Order Logic-Based Method for Enhancing Logical Reasoning Capabilities of LLMs. In *NeurIPS Workshop on FoRLM*.

Morishita, T.; Morio, G.; Yamaguchi, A.; and Sogawa, Y. 2024. Enhancing reasoning capabilities of llms via principled synthetic logic corpus. In *NeurIPS*.

Wang, X.; Yang, H.; Cheng, F.; and Liu, F. 2026. Adaptive Selection of Symbolic Languages for Improving LLM Logical Reasoning. In *AAAI Workshop on P-AI-FM*.

Xu, J.; Fei, H.; Pan, L.; Liu, Q.; Lee, M.-L.; and Hsu, W. 2024. Faithful Logical Reasoning via Symbolic Chain-of-Thought. In *ACL*.

Ye, X.; Chen, Q.; Dillig, I.; and Durrett, G. 2023. Satlm: Satisfiability-aided language models using declarative prompting. In *NeurIPS*.