



Symbol-LLM: Symbolic Knowledge Injection into LLMs

Task 5 – Understanding Logical Reasoning

Haardhik Mudagere Anil

Date: May 2025



What Is the Problem?

LLMs are great at natural language, but struggle with:

- Symbolic logic (e.g., Prolog, FOL)
- Precise multi-step reasoning

Goal of Symbol-LLM:

- Enable LLMs to reason over symbolic data while retaining language skills

What is Symbolic Data?

- Symbolic data refers to information represented in a formal, logic-based, discrete structure - not natural language.
- Examples:
 - Prolog: `parent(homer, bart).`
 - FOL: $\forall x \text{ Student}(x) \rightarrow \text{Enrolled}(x, \text{Math})$
 - SQL, API calls, math expressions
 - Programming Instructions (C++, Python)
- LLMs need to understand symbolic forms to reason logically

What Symbol-LLM Proposes

Two-stage training pipeline:

- 1. Injection - Train on symbolic data
- 2. Infusion - Combine symbolic + NL data

Result:

- Understands symbolic formats
- Retains general NLP capabilities
- Generalize across symbol types (Prolog, FOL, SQL, etc.)

Evaluation Summary

- Tested on 34 symbolic tasks + general NLP benchmarks
- Outperforms GPT-3.5, Claude, LLaMA on:
 - FOL translation
 - SQL
 - API calls
 - AMR parsing
- Maintains NLP performance while improving symbolic reasoning
- Symbol-LLM learns symbols + language = robust reasoning

My Experiment - Simpsons KB

- Used Simpsons Prolog KB with facts like `parent(homer, bart)`.
- Converted facts to natural language:
- "Homer is Bart's parent."
- Prompted GPT-4 with question:
- Who are Bart's grandparents?
- LLM Output: Abe and Mona - correct reasoning!

Takeaways

- Symbol-LLM works with simple prompts
- GPT-4 can answer symbolic queries if injected properly
- Easy to test Prolog KB using natural language

References

- Symbol-LLM Paper: <https://arxiv.org/abs/2311.09278>
- Simpsons KB from Task 3
- Prolog for symbolic ground truth