SPARK: Step-by-step Proof Assistant for Reasoning and Knowledge

Language Warriors

Srivatsa   **(2024701003)**
Aditya   **(2021114009)**

Motivation

Mathematical reasoning is a cornerstone of scientific and technological advancement, yet existing AI systems often struggle with rigorous, step-by-step problem-solving and verification. Even after giving questions we majorly receive wrong answers.

This research bridges the gap by developing an interactive, math assistant that combines LLM reasoning with formal verification.

- By enabling step-by-step explanations, auto-verification, and cross-domain adaptability, this system empowers us to tackle complex problems with confidence.
- It also advances AI fairness and interpretability, ensuring reliable and ethical mathematical reasoning tools for diverse applications.
- We also aim to democratise the reasoning capabilities of LLM by experimenting within the limits of resource by working on small LLMs and consumer grade GPU.

Literature Survey

Direct Preference Optimization (DPO):

- Rafailov, R., Sharma, A., Mitchell, E., Ermon, S., Manning, C. D., & Finn, C. (2023). "Direct Preference Optimization: Your Language Model is Secretly a Reward Model." arXiv preprint arXiv:2305.18290.
  arXiv

rStar-Math: Small LLMs Can Master Math Reasoning with Self-Evolved Deep Thinking

- Guan, X., Zhang, L. L., Liu, Y., Shang, N., Sun, Y., Zhu, Y., Yang, F., & Yang, M. (2025). *rStar-Math: Small LLMs Can Master Math Reasoning with Self-Evolved Deep Thinking*. arXiv preprint arXiv:2501.04519. Retrieved from https://arxiv.org/abs/2501.04519

ReAct Framework:

- Yao, S., Zhao, J., Yu, D., Cao, Y., Yu, Z., & Huang, D. (2022). "ReAct: Synergizing Reasoning and Acting in Language Models." arXiv preprint arXiv:2210.03629.

Proposal

This project develops an interactive AI assistant for mathematical reasoning, combining LLMs/SLMs with formal verification targeting low resource systems and low cost training paradigms. It enables step-by-step problem-solving, auto-verification, and cross-domain adaptability. By integrating methods like rStar, DPO, and ReAct, the system ensures reliable, interpretable, and reasoning, advancing AI fairness which can then be used in complex mathematical tasks.

Datasets

1. GSM8K (7.4k points)
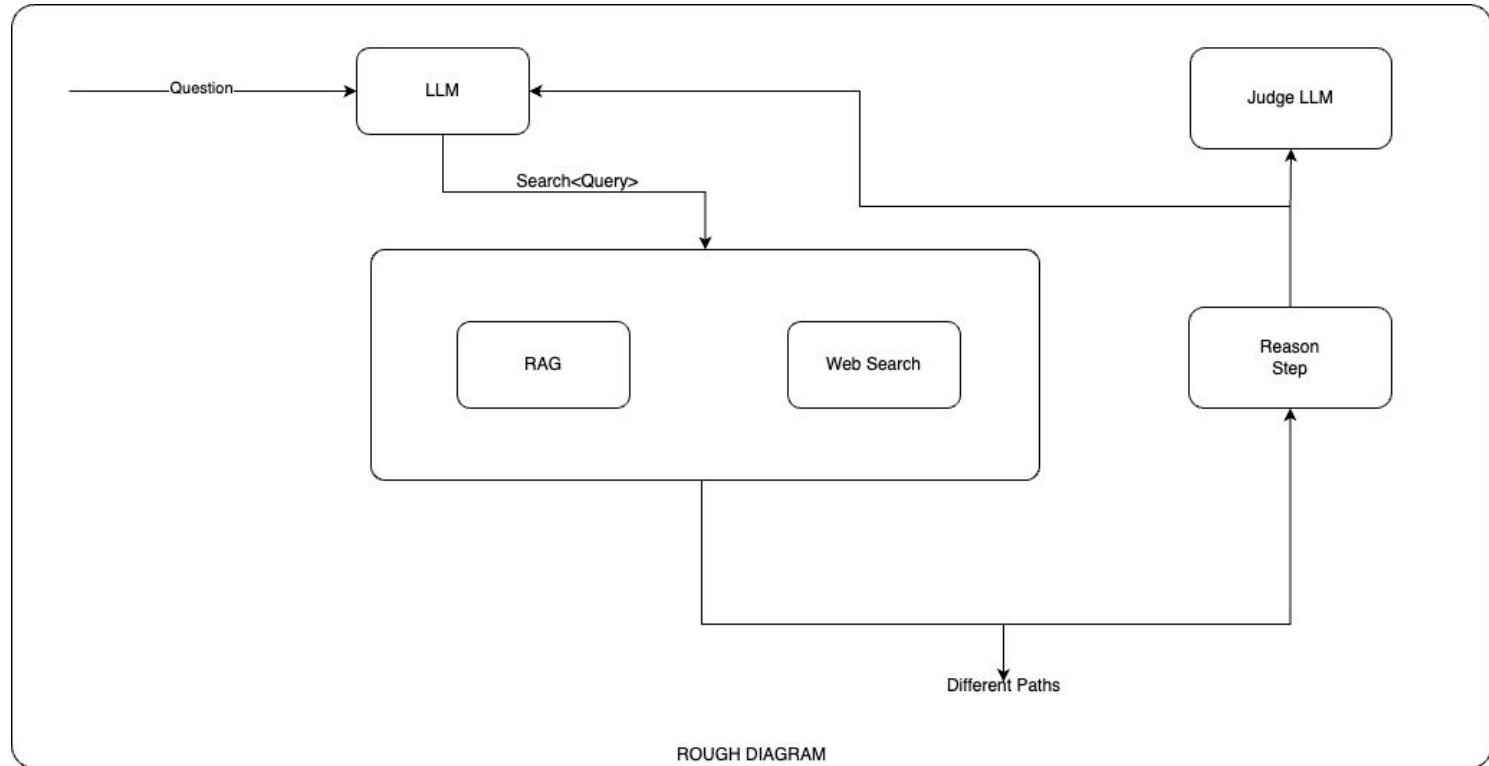2. AI-MO/ NuminaMath-Cot
3. MATH (~50K)

Experiments

1. Baseline :
   a. Zero and few shot prompting
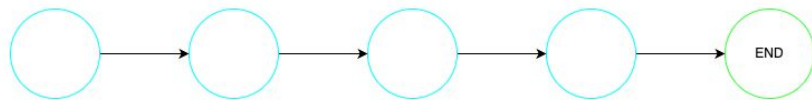2. Multi-step reason generation using theory augmentation assisted by judge LLMs.
   a. Linear step Generation
   b. Tree step Generation
3. Explore RL based and RL-free methods (stretched goal)
   a. rStar framework - RL with AI feedback (PPO)
   b. DPO   - Direct Preference Optimization

Note :

- We will make use of GPU efficient fine tuning of quantized SLMs
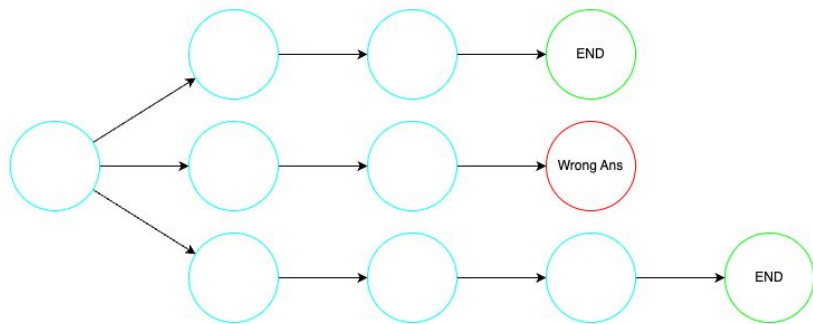
Diagrams

Question → LLM

Search<Query>

RAG   Web Search

Judge LLM

Reason Step

Different Paths

ROUGH DIAGRAM

LINEAR STEP
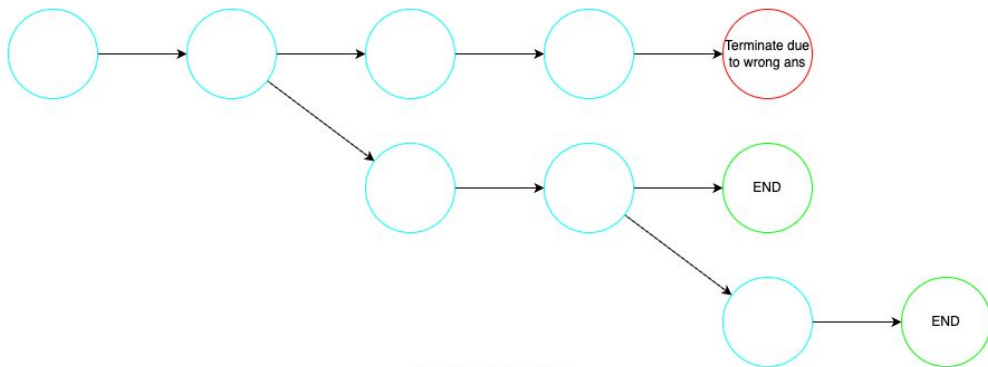
TREE LIKE MULTI-STEP

Terminate due to wrong ans

END

END

TREE LIKE MULTI-STEP

END

Wrong Ans

END

Targeted Conference

- ACL 2025 Student Research Workshop (SRW)
  - Pre-submission mentoring deadline: March 27, 2025
  - Pre-submission feedback: May 1, 2025
  - Paper submission deadline: May 18, 2025
  - Review deadline: June 6, 2025
  - Acceptance notifications: June 21, 2025
  - Camera-ready deadline: July 1, 2025
  - ACL 2023 conference dates: July 28-30, 2025

Expected Timeline

Date	Work
3rd Week of Feb	Evaluation pipelines + Baseline Implementations
1st March	Linear Step + Tree Step generation
14th March	RL based and RL-free methods
21st March	Evaluation Metrics and start of Paper writing

Thank You