

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)

Dataset Viewer		Auto-converted to Parquet	API	Embed	Full Screen Viewer
Subset (2)		Split (2)			
main · 8.79k rows		train · 7.47k rows			
Search this dataset		SQL Console			
question string · lengths		answer string · lengths			
 137~232 46%		 50~168 19.9%			
Natalia sold clips to 48 of her friends in April, and then she sold half as many clips in May. How many...		Natalia sold 48/2 = <<48/2=24>>24 clips in May. Natalia sold 48+24 = <<48+24=72>>72 clips altogether...			
Weng earns \$12 an hour for babysitting. Yesterday, she just did 50 minutes of babysitting. How much did she...		Weng earns 12/60 = <<12/60=0.2>>0.2 per minute. Working 50 minutes, she earned 0.2 x 50 =...			
Betty is saving money for a new wallet which costs \$100. Betty has only half of the money she needs. Her...		In the beginning, Betty has only 100 / 2 = <<100/2=50>>50. Betty's grandparents gave her 15 + 2...			

Dataset Viewer

Auto-converted to Parquet

API

Embed

Full Screen Viewer

Split (1)

train · 50k rows

Search this dataset

SQL Console

role_1 string · classes	topic; string · classes	sub_topic string · lengths	message_1 string · lengths	message_2 string · lengths
<div><div></div><div>1 value</div></div>	<div><div></div><div>25 values</div></div>	<div><div></div><div>9190</div></div>	<div><div></div><div>213.58k</div></div>	<div><div></div><div>16.33k</div></div>
Mathematician_RoleType.ASSISTANT	Algebra	Solving linear equations	Solve for x: 5x + 3 = 28	To solve for x, we need to isolate x...
Mathematician_RoleType.ASSISTANT	Algebra	Solving linear equations	Solve for x: 2x + 5 = 9- x.	To solve for x, we will follow these...
Mathematician_RoleType.ASSISTANT	Algebra	Solving linear equations	Solve the equation: 3x + 5 ...	To solve the equation 3x + 5 = ...

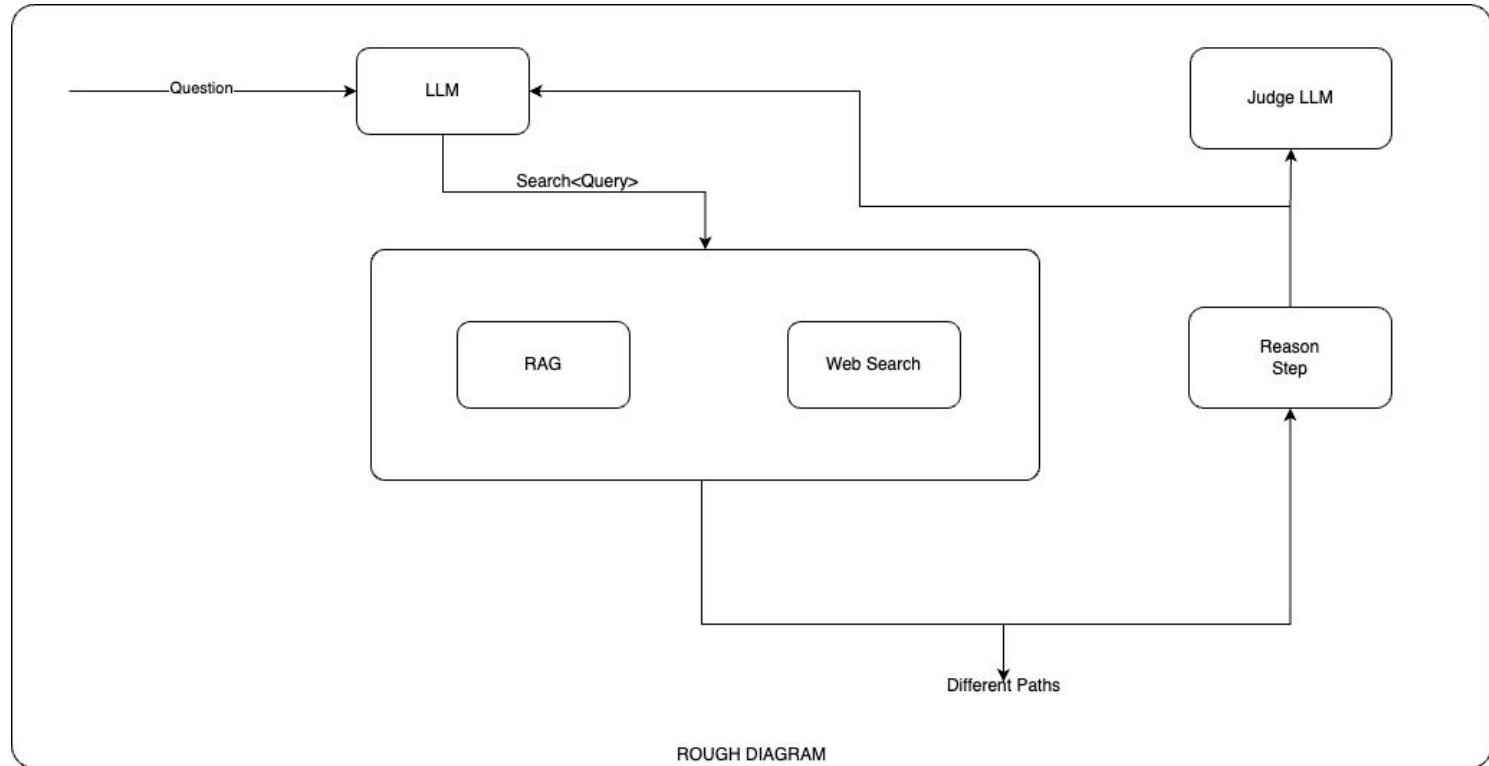
Experiments

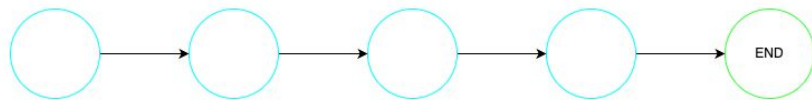
1. Baseline :
 - a. Zero and few shot prompting
2. Multi-step reason generation using theory augmentation assisted by judge LLMs.
 - a. Linear step Generation
 - b. Tree step Generation
3. Explore RL based and RL-free methods (stretched goal)
 - a. rStar framework - RL with AI feedback (PPO)
 - b. DPO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

Note :

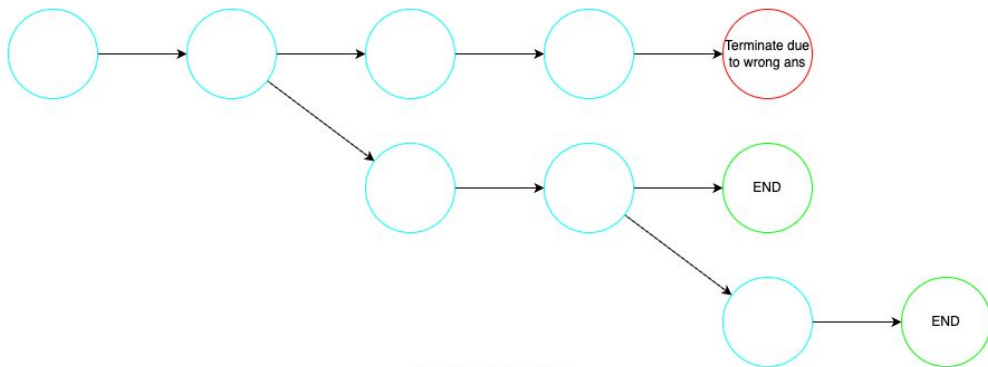
- We will make use of GPU efficient fine tuning of quantized SLMs

Diagrams

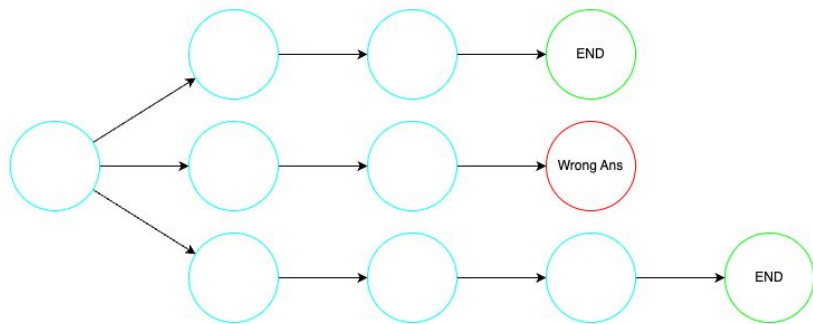




LINEAR STEP



TREE LIKE MULTI-STEP



TREE LIKE MULTI-STEP

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