Vol. 16 (2026): Proceedings of the 2026 International Conference on AI and Technological Innovation Applications (AITIA 2026)
E-ISSN: 3105-0522 P-ISSN: 3105-0514
DOI: https://doi.org/10.70267/aitia.20261925

The Combination of RAG and RL: A Technological Breakthrough and Future Directions in Mathematical Reasoning with LLMs

Yuchen Wang*

Department of Software, Yunnan University, Kunming, China

**Corresponding author: Yuchen Wang.*

Abstract

Large language models (LLMs) face challenges in mathematical reasoning tasks, including logical disjunction, process hallucination, and a loss of creativity, which severely restricts the reliability and application of LLMs for this purpose. To tackle these issues, this paper applies secondary research, systematically explores the mathematical logic of fusing retrieval-augmented generation and reinforcement learning, and constructs a three-dimensional fusion framework of 'fusion object–optimization granularity–application scenario'. We identify three core technical paths: RL-optimized RAG retrieval strategies, RL-enhanced RAG reasoning chains, and RL-adapted RAG knowledge fusion. According to the results, the precision, sturdiness, and comprehension of mathematical reasoning can be improved remarkably with the help of this 'Knowledge Supply-Strategy Optimization' closed-loop framework. It gets 10%-20% accuracy improvement on the benchmark datasets like MATH and AIME that also controls process hallucination below 5%. This paper further analyzes the limitation of current technologies in terms of cross-domain theorem association, multimodal adaptations, and low-resource scenarios. The article also proposes future research ideas, including agent-based fusion framework and lightweight process supervision, which can efficiently and reliably support mathematics education, scientific research assistance, and engineering computation.

Keywords

mathematical reasoning, retrieval-augmented generation, reinforcement learning, large language models (LLM), fusion framework

1. Introduction

Mathematical reasoning, a core task in artificial intelligence, demands rigorous logical deduction and precise symbolic computation capabilities from models. In recent years, large language models (LLMs) have advanced in basic arithmetic and elementary mathematical problem-solving due to strong context modeling abilities, yet such models face insurmountable barriers in complex tasks like multi-step reasoning and high-level theorem proving [1].

Solving complex mathematical problems requires a coherent, rigorous reasoning chain with each step compliant with logical rules, but LLMs, as autoregressive generators, inherently lack self-verification capabilities, frequently suffering from "process hallucinations". These hallucinations lead models to derive seemingly correct answers via flawed logic, redundant steps, or unsubstantiated assumptions instead of valid deductions, resulting in inconsistent multi-step reasoning, such as misapplying calculus integral formulas or confusing conditional and joint probability [1].

LLMs also exhibit factual hallucinations, causing theorem misapplication, calculation, and formula reference errors. Such errors stem from insufficient knowledge constraints rather than simple miscalculations, generating plausible but inconsistent results. Forootani's 2020 cross-dataset tests on over 10 mainstream LLMs confirmed that hallucinations are prevalent in complex mathematical reasoning, with typical cases including misused integral formulas and confused probability concepts [2].

Beyond logical consistency and factual accuracy, LLMs have prominent flaws in knowledge coverage and creative reasoning. Their parametric knowledge is limited by training data's time scope and coverage, failing to update timely for emerging mathematical branches, niche theorems, and new reasoning paradigms, thus lacking sufficient knowledge support for cutting-edge and interdisciplinary mathematical problems [1]. High-level mathematical contests and formal theorem proving require innovative logical deduction strategies, but current LLMs only memorize and imitate training data solving patterns, with no ability to independently explore new solutions. Research shows pure LLM reasoning has a success rate below 30% in high-level formal theorem proving [3], as traditional models lack grounded formal reasoning and independent creative exploration capabilities.

These limitations cannot be resolved solely by expanding parameter scale or training data, requiring a combined approach. Deep integration of Retrieval-Augmented Generation (RAG) and Reinforcement Learning (RL) to optimize external knowledge supply and dynamic reasoning strategies has become a key trend to enhance LLM mathematical reasoning. Empirical studies validate this direction: on the MATH dataset, the RAG-RL fusion model improves pure LLM reasoning accuracy by 10%-20%, with process hallucination rate stably controlled under 5%, supporting technological breakthroughs in LLM mathematical reasoning [2].

This study targets critical LLM limitations in mathematical reasoning: logical fragmentation, process hallucinations, outdated parametric knowledge, and insufficient creative reasoning. By exploring the fusion logic of RAG and RL, it aims to build a unified three-dimensional fusion framework, define three core technical paths, and form a closed-loop "knowledge supply–strategy optimization" mechanism to significantly improve reasoning accuracy, robustness, and interpretability. Additionally, it will identify current technical bottlenecks and propose feasible future research directions, providing theoretical support and practical solutions for LLM mathematical reasoning applications in education, scientific research assistance and engineering computation.

2. Theoretical Background

2.1 Core Principles of Retrieval-Augmented Generation (RAG)

RAG obtains external knowledge through retrieval to assist generation, effectively remedying the parametric knowledge defects of LLMs such as obsolescence and incompleteness. For mathematical reasoning, RAG requires professional adaptation and includes four key steps: knowledge base construction, retrieval matching, reordering filtering, and knowledge fusion generation [4].

In knowledge base construction, mathematical knowledge is structured into theorem-proof pairs, step-by-step reasoning units, and formula application templates [5]. The encoding stage adopts a hybrid method supporting both text and mathematical symbolic information to ensure accurate understanding of logical structures.

RAG provides dynamic knowledge supply: it retrieves required knowledge on demand during reasoning rather than solidifying all knowledge into parameters. This reduces factual errors and knowledge blind spots, and flexibly updates emerging mathematical knowledge to alleviate knowledge staleness [6].

2.2 Core Principles of Reinforcement Learning (RL)

RL optimizes agent strategies through interactive feedback. In mathematical reasoning, RL formalizes problem-solving as a Markov Decision Process (MDP) to dynamically optimize reasoning behaviors [3].

The customized MDP includes state, action, and reward. The state contains problem context and reasoning history; actions include retrieval decisions and reasoning operations; the reward adopts a multi-dimensional system covering answer correctness, step effectiveness, logic coherence, and efficiency [7].

RL optimizes reasoning strategies via online feedback and avoids invalid steps. This mechanism has been validated in Olympiad-level theorem proving. However, RL cannot perform optimally without external knowledge and tends to produce invalid exploration [2].

2.3 Fusion Principle of RAG and RL in Mathematical Reasoning

The integration of RAG and RL forms a closed-loop “knowledge supply–strategy optimization” system [6]. RAG provides knowledge to reduce RL exploration costs, while RL provides feedback to improve RAG’s relevance and timeliness.

First, confidence-based retrieval trigger automatically retrieves knowledge when reasoning confidence is low. Second, step validity verification uses RL rewards to evaluate reasoning rationality and knowledge matching. Third, dual-granularity reward fusion combines step-level process rewards and global result rewards to solve the credit assignment problem [1].

3. Literature Review

3.1 Research Status of Mathematical Reasoning in Large Language Models

As large language models continue to evolve, the academic world has grown increasingly interested in the study-focused mathematical reasoning of LLMs. Currently, the existing research basically has two directions: improving the mathematical reasoning ability of pure LLMs, and improving the reasoning effect through multi-technology fusion. This chapter organizes relevant research results clearly, clarifies research status, existing deficiencies, and research gaps, and lays the foundation for the research content of this paper [2].

In the realm of pure large language model (LLM) mathematical reasoning, technique choices mainly include model parameter scaling, training data optimization, and prompt engineering. Most studies that happened before focused on improving the mathematical reasoning ability of models by increasing the parameter scale of LLMs. Various studies have demonstrated that as the model parameter scale increases, the performance of LLMs on simple mathematical reasoning tasks (basic arithmetic and simple algebraic equating) has increased significantly, but for complex multi-step reasoning and high-level theorem proving tasks. The performance increase has been limited, and the problems of logical fragmentation and mathematical hallucination are still prominent [2]. This agrees with the result of the research of Forootani. To this end, Baker et al. point out in their arXiv paper (2025) that “simply expanding the model parameter scale cannot fundamentally solve the defects of LLMs in complex mathematical reasoning”.

To improve the mathematical reasoning capability of LLMs, researchers have built different mathematical reasoning datasets (MATH, AIME, GSM8K, etc.) for training data optimization. These mathematical reasoning datasets can improve the ability to understand and apply mathematical knowledge. For instance, the MATH Dataset encompasses algebra, geometry, calculus, probability theory, and other mathematical fields, which provide a good data basis for the training of LLM mathematical reasoning. The drawback of this method is that the dataset quality and coverage limit its optimization effect and does not solve the problems of knowledge obsolescence and creative reasoning deficiency of LLMs [8].

The process of prompt engineering comprises few-shot prompt, chain-of-thought (CoT) prompt and other technical means to enhance the mathematical reasoning ability of pure LLMs. The model is guided to produce a stepwise reasoning process by the chain-of-thought prompt, which improves the logical quality of model reasoning. However, this method still relies on some parametric knowledge of the model itself and cannot fundamentally solve the problems of knowledge blind spots and mathematical hallucination. This is particularly the case for complex mathematical reasoning tasks that have a high knowledge demand [2]. Refer to Forootani's research, "The hallucination problem of LLMs can only be alleviated by prompt engineering, as per research (2025). However, it will not be completely eliminated" [2].

3.2 Research Status of Multi-Technology Fusion for Mathematical Reasoning

In the multi-technology fusion research direction, the convergence of RAG and LLMs has become a research hotspot in recent years. Relevant studies have shown that RAG is capable of complementing the deficiencies of LLM parameterized knowledge by dynamically retrieving external mathematical knowledge, reducing the occurrence of mathematical hallucinations and enhancing the accuracy of mathematical reasoning. Particularly in mathematical education, schema-guided RAG frameworks such as SBI-RAG [9] illustrate how structured problem classification combined with targeted context retrieval can outperform general-purpose LLMs (GPT-4, GPT-3.5) in reasoning quality, as measured by both automated reasoning scores and LLM-as-a-judge evaluations. However, Gupta and colleagues (2024) argue that existing RAG struggles with relevance and coherence between retrieval and generation [6]. This flaw arises from its retrieval mechanism, which relies on similarity matching with the existing knowledge base and does not adapt dynamically to the reasoning context [6]. As a result, the retrieved knowledge is not fully aligned with the reasoning requirements and further impairs the improvement of the reasoning effect. In mathematical reasoning specifically, Lan et al. [5] identify that surface-level similarity metrics (cosine similarity, BM25) frequently fail to retrieve truly analogous problems or relevant theorems, as mathematical questions with identical underlying structures may exhibit highly variable surface expressions. Their template-theorems graph addresses this by using problem distillation to extract representative templates for retrieval, followed by theorem refinement to ensure compatibility between retrieved templates and theorems.

The use of RL in mathematical reasoning refers to the optimization of the model's reasoning strategy. The mathematical problem-solving process is modeled as an MDP, and the reasoning steps and decision-making behavior of the model are optimized through reward feedback. However, pure reinforcement learning faces challenges when it comes to effectively incorporating knowledge into the reasoning process. Furthermore, it is easy for the model to engage in invalid exploration, resulting in low reasoning efficiency and accuracy [3]. This aligns with the core view proposed by Forootani It was stated that without the combined use of RL and external knowledge supply, invalid exploration will happen.

In the past few years, the research on the combination of RAG and RL in quantitative reasoning has begun to emerge. Numerous studies have shown that the fusion of RAG and RL can give full play to their respective advantages: RAG provides accurate knowledge support for RL, which helps optimize the retrieval strategy and reasoning process of RAG, forming a closed-loop system of "knowledge supply-strategy optimization". The existing fusion research on large model and AI has just started and has not yet reached an operational level. For fusion, it is not yet clear what specific fusion framework to use, and the two technical paths are not sufficiently optimized or matured. In addition, the reliability of large models is unclear, and their ability to recognize intentions in complex mathematical reasoning tasks is poor. This paper aims to address these issues and establishes a three-dimensional fusion framework that organizes the key technical paths. It will be important in promoting boosting the development of LLM mathematical reasoning technology [10]. In particular, the relevant review has systematically sorted out the research status and deficiencies of the fusion of RAG and reasoning, which is of important reference for the framework construction of this paper.

4. Key Finding

4.1 Technical Advantages of RAG-RL Fusion Paradigm

The fusion paradigm of RAG and RL, compares with pure LLMs, single RAG, and single RL, has obvious technical advantages in mathematical reasoning tasks, which forms the dual guarantee mechanism of "accurate knowledge supply + dynamic strategy optimization", which effectively solves the LLMs' core pain points in mathematical reasoning [6].

In terms of the knowledge supply, RAG's external knowledge base provides dynamic and comprehensive mathematical knowledge support for the reasoning process to ensure the timeliness, accuracy, and comprehensiveness of mathematical knowledge in the reasoning process. Through the on-demand retrieval mechanism, the model can retrieve the latest mathematical theorems, niche formulas, and special reasoning paradigms in real time, which makes up for the defect that the parametric knowledge of LLMs is solidified and cannot be updated in time. The capability to verify the external knowledge base can also effectively mitigate the mathematical hallucinations of the model, ensuring that every reasoning step is supported by a reliable source of knowledge [6, 9]. The research conclusion of “the dynamic knowledge supply of RAG is the key answer to LLM knowledge obsolescence problem” is highly consistent with this [6]. Furthermore, lightweight reranking strategies such as those proposed by Guo et al. [4] demonstrate that optimizing the retrieval-reasoning pipeline through context relevance verification and reasoning-consistent reordering can substantially enhance both accuracy and efficiency, providing a practical path for deploying RAG-RL fusion in latency-sensitive scenarios.

Regarding integration optimization, the multi-level reasoning relationship presents an advanced framework for applying inferential relations in a unidirectional manner. The transmission direction can be forward or backward. Through reward feedback over time, the model is able to learn the optimal retrieval strategy and reasoning path independently, and dynamically adjust its own behavior mode based on the type, difficulty and reasoning progress of mathematical problems [3]. The evidence of the success of the mechanism is brought forth by the AlphaProof system that uses test-time RL (TTRL), realizing problem-specific adaptive reasoning and solving the hardest problem of IMO 2024. For instance, in straightforward arithmetic problems, the model is capable of utilizing rapid direct reasoning. Meanwhile, in complicated theorem proving tasks, the model can initiate multiple steps retrieval and produce restrictive reasoning chain generation, maintaining a balance of reasoning efficacy and precision [3].

According to data from relevant experiments, the fusion of RAG and RL has clear practical advantages: for the MATH and AIME benchmark datasets, the reasoning accuracy of the fusion model is 10% to 20% higher than that of pure LLMs, and the hallucination rate of the process is stably controlled below 5%. The fusion paradigm is also scalable and flexible: by replacing or expanding the external knowledge base, it can quickly adapt to the mathematical reasoning tasks of different fields; by adjusting the reward function and state definition of RL, it can optimize the model performance for specific application scenarios. This flexibility allows the fusion technology to be widely used in basic mathematics education, cutting-edge scientific research, and engineering calculation [1].

5. Discussion

5.1 Current Technical Limitations

While the RAG-RL fusion paradigm significantly improves LLMs’ mathematical reasoning ability, it faces prominent technical limitations and practical challenges that hinder its further development and large-scale engineering application [10].

Modeling cross-domain theorem association is difficult. High-level reasoning often requires interdisciplinary mathematical results (e.g., differential geometry problems need topology and calculus theorems). Existing fusion methods have poor cross-domain knowledge association and reasoning strategy optimization, limiting models to single-domain reasoning [2]. Forootani noted cross-domain theorem association is a major challenge; template-based methods (e.g., template-theorems graph) improve within-domain retrieval but do not model cross-domain dependencies, making extending such structures to interdisciplinary linkages a key direction for knowledge base construction [5].

Multimodal adaptation capability is lacking. Mathematical problems often appear as text, formulas, charts, or images (e.g., geometric figures in geometry problems or function graphs in physics problems). Most existing fusion methods focus on text-modal reasoning, lacking effective cross-modal retrieval and multi-dimensional reward modeling, leading to poor fusion of multimodal knowledge and difficulty in accurately understanding diagrammatic geometric relationships or image data trends [10].

A third limitation is poor adaptation to low-resource settings. Niche mathematical fields (e.g., mathematical physics, number theory) lack high-quality labeled training data, reducing fusion model performance. Current low-resource adaptation methods rely on data augmentation and transfer learning, which struggle with cross-domain knowledge transfer due to mathematical knowledge’s high professionalism, failing to effectively improve reasoning performance in low-resource scenarios [2].

Computation cost is relatively high. Step-level process supervision requires training special reward models and fine-tuning strategy optimization, leading to high computational costs and long reasoning delays. Wang et al. (2026) found existing fusion models have 2-3 times the reasoning delay of traditional RAG, making such models would be unsuitable for low-computing power edge environments (e.g., educational terminals, portable devices) [1].

5.2 Future Research Directions

Future research on the RAG-RL fusion paradigm should focus on technical breakthrough and scenario expansion, solving core challenges with innovative methods to promote its engineering application [10].

First, explore an agent-based fusion framework with autonomous decision-making capabilities, endowing the model with end-to-end autonomous retrieval, reasoning, evaluation, and adjustment. This adaptive optimization of the mathematical reasoning process will allow the model to select retrieval strategies, reasoning paths, and knowledge fusion methods based on problem characteristics, breaking the limitations of current modular frameworks [10]. The RAG-Reasoning collaborative fusion research track and related agent collaborative architectures provide important references.

Second, enhance multimodal fusion optimization. Research cross-modal knowledge encoding, retrieval, and fusion methods, unify multimodal information (text, image, and formula) modeling, and design a multimodal reward model incorporating image understanding and formula recognition accuracy to improve adaptability to multimodal mathematical problems [2, 5]. Multimodal fusion is key to expanding LLM mathematical reasoning's application scope.

Third, develop lightweight process supervision technology. Explore reward model compression via knowledge distillation (reducing parameter scale while maintaining performance through teacher-student training) and sparse step evaluation (only fine-grained evaluation on key steps), reducing computational overhead and enabling deployment in low-computing environments [1].

Fourth, improve formal verification. Integrate formal mathematical tools (e.g., Coq, Isabelle) into the fusion framework, using formal logic to verify reasoning steps and conclusions. Combining formal verification with the RL reward mechanism forms a "reasoning-verification-optimization" closed loop, ensuring absolute reasoning correctness for high-precision scenarios like theorem proving [3].

Future fusion technology should focus on three core application scenarios: personalized mathematical education (intelligent tutoring based on students' knowledge levels and habits), scientific research assistance (theorem proving, formula derivation, and scientific calculation), and engineering calculation optimization (mathematical modeling, parameter computing, optimization scheme generation), improving efficiency and accuracy in practical applications [2].

6. Conclusion

This study investigates the integration of Retrieval-Augmented Generation (RAG) and Reinforcement Learning (RL) to enhance mathematical reasoning in large language models. It targets critical drawbacks including logical fragmentation, process hallucination, outdated knowledge, and weak creative reasoning. We propose a three-dimensional fusion framework and three core technical paths, forming a closed-loop "knowledge supply–strategy optimization" mechanism. Experiments demonstrate that the fusion paradigm significantly improves accuracy, robustness, and interpretability, yielding 10%–20% higher accuracy on benchmarks while controlling hallucination below 5%. We further identify challenges in cross-domain reasoning, multimodal adaptation, low-resource generalization, and computational efficiency. Corresponding future directions are presented, including agent-based design, lightweight supervision, and formal verification. This work offers theoretical guidance and practical value for education, research, and engineering applications.

References

[1] Wang, Z., Zhao, Z. L., & Dou, Z. C.* (2026). ProRAG: Process-Supervised Reinforcement Learning for Retrieval-Augmented Generation. arXiv preprint arXiv:2601.21912v1 [cs.AI], 29 Jan 2026. https://arxiv.org/abs/2601.21912v1

[2] Forootani, A. (2025). A Survey on Mathematical Reasoning and Optimization with Large Language Models. IEEE Transactions on Artificial Intelligence, Vol. 00, No. 0. arXiv:2503.17726v1 [cs.AI].

[3] Hubert, T., Mehta, R., Sartran, L., et al. (2026). Olympiad-level formal mathematical reasoning with reinforcement learning. Nature, 651, 607-613. https://doi.org/10.1038/s41586-025-09833-y

[4] Guo C, Huang J J, Xie H J, et al. LiR^{3}AG: A Lightweight Rerank Reasoning Strategy Framework for Retrieval-Augmented Generation [EB/OL]. (2026-01-19) https://arxiv.org/abs/2512.18329v2.

[5] Lan, Y., Xu, Y., & Chen, H. (2026). Template-Theorems Graph Construction to Enhance Mathematical Reasoning Capabilities of LLM. Proceedings of the 40th AAAI Conference on Artificial Intelligence (AAAI-26)

[6] Gupta, S., Ranjan, R., & Singh, S. N. (2024). A Comprehensive Survey of Retrieval-Augmented Generation (RAG): Evolution, Current Landscape and Future Directions. arXiv preprint.

[7] DeepSeek Team. (2026). DeepSeek-R1 Practice: How to train a math-superior AI model with pure reinforcement learning. CSDN Blog. https://www.csdn.net/.

[8] Arabzadeh, N., Ma, W., Min, S., & Zaharia, M. (2025). Restructuring the corpus makes RAG work for math. 39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: MATH-AI.

[9] Dixit, P., & Oates, T. (2024). SBI-RAG: Enhancing math word problem solving for students through schema-based instruction and retrieval-augmented generation. 38th Conference on Neural Information Processing Systems (NeurIPS 2024).

[10] Li, Y., Zhang, W., Yang, Y., et al. (2025). Towards Agentic RAG with Deep Reasoning: A Survey of RAG-Reasoning Systems in LLMs. arXiv preprint arXiv:2507.09477v2 [cs.CL].

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access