FlowKV Impact on Reasoning Performance in RAG-Enhanced LLMs for Long Conversations

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 8 peer-reviewed papers addressing the following research question: Does FlowKV improve reasoning performance in RAG-enhanced LLMs on benchmarks like MMLU or GSM8K when handling long, multi-turn problem-solving conversations (e.g., 50+ turns). 13 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 2.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Building Math Agents with Multi-Turn Iterative Preference Learning. Research question: Does FlowKV improve reasoning performance in RAG-enhanced LLMs on benchmarks like MMLU or GSM8K when handling long, multi-turn problem-solving conversations (e.g., 50+ turns)?.

2 Methodology

Systematic literature search across multiple databases yielded 8 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 2.8/10.

3 Results

8 papers retrieved. 13 claims extracted; 0 independently verified. Quality review score: 2.8/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
The MATH dataset includes 5K problems across diverse mathematical fields such as algebra, geometry, probability, number	×	0.03
The GSM8K test set consists of 1319 grade-school math word problems.	×	0.02
The augmented prompt set includes 7.5K training problems from MATH and 7.47K training problems from GSM8K.	×	0.08
The final training set consists of 60K training prompts.	×	0.02
The maximal number of rounds H is set to 6.	×	0.01
Gemma-1.1-it-7B M-DPO Iteration 3 achieves 83.9 on MATH, 51.2 on GSM8K, and 67.6 overall.	×	0.08
Gemma-1.1-it-7B Iterative M-KTO achieves 82.1 on MATH, 49.5 on GSM8K, and 65.8 overall.	×	0.06
CodeGemma-1.1-it-7B Iterative M-DPO achieves 81.5 on MATH, 50.1 on GSM8K, and 65.8 overall.	×	0.03
CodeGemma-1.1-it-7B Iterative M-KTO achieves 81.6 on MATH, 49.6 on GSM8K, and 65.6 overall.	×	0.04
Mistral-7B-v0.3 Iterative M-DPO achieves 82.3 on MATH, 47.5 on GSM8K, and 64.9 overall.	×	0.03
Mistral-7B-v0.3 Iterative M-KTO achieves 81.7 on MATH, 46.7 on GSM8K, and 64.2 overall.	×	0.05
Gemma-2-it-9B Iterative M-DPO achieves 86.3 on MATH, 54.5 on GSM8K, and 70.4 overall.	×	0.09
Gemma-2-it-9B Iterative M-KTO achieves 86.1 on MATH, 54.5 on GSM8K, and 70.3 overall.	×	0.09

References

- http://arxiv.org/abs/2601.06757v1
- http://arxiv.org/abs/2312.17080v4
- http://arxiv.org/abs/2409.02392v2