

Multi-Turn Iterative Preference Learning vs. Supervised Fine-Tuning for Zero-Shot Mathematical Reasoning

Assignee Research

June 1, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: How does the multi-turn iterative preference learning approach compare to Supervised Fine-Tuning (SFT) in improving zero-shot generalization of LLMs on mathematical reasoning tasks, as measured by. Recent studies have shown that large language models' (LLMs) mathematical problem-solving capabilities can be enhanced by integrating external tools, such as code interpreters, and employing multi-turn Chain-of-Thought (CoT) reasoning. While current methods focus on synthetic. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.2/10. 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: How does the multi-turn iterative preference learning approach compare to Supervised Fine-Tuning (SFT) in improving zero-shot generalization of LLMs on mathematical reasoning tasks, as measured by accuracy on the GSM8K benchmark?.

2 Methodology

Systematic literature search across multiple databases yielded 15 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.2/10.

3 Results

15 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.2/10.

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.

References

- <http://arxiv.org/abs/2409.02392v2>
- <http://arxiv.org/abs/2602.07464v1>
- <http://arxiv.org/abs/2406.18629v1>