Training Strategies for Language Model Generalization in Mathematical Reasoning

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: What training strategies improve language model generalization to novel mathematical reasoning problems v19. 12 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Improving Large Language Model Fine-tuning for Solving Math Problems. Research question: What training strategies improve language model generalization to novel mathematical reasoning problems v19.

2 Methodology

Systematic literature search across multiple databases yielded 16 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.7/10.

3 Results

16 papers retrieved. 12 claims extracted; 0 independently verified. Quality review score: 4.7/10.

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 is used for experiments, with 4.5K original test examples for training and validation, and 500 test exa	×	0.05
Two sources of correct step-by-step solutions are used for model training: original human-written explanations in the MA	×	0.13
The automatic grading script provided by Lightman et al. (2023) checks the mathematical equivalence of generated solutio	×	0.03
Greedy decoding is used for Pass@1 performance evaluation, and nucleus sampling is used for majority voting performance	×	0.10
The sampling temperature is set to 0.6, and the top-p value is set to 0.95 for nucleus sampling.	×	0.01
PaLM 2-S* and PaLM 2-L are fine-tuned on step-by-step solutions with the MLE training objective.	×	0.12
Three fine-tuning strategies are explored: using original MATH solutions only, using PRM800K GPT-4 solutions only, and u	×	0.10
Fine-tuning is generally helpful for achieving better performance compared to the few-shot performance of pre-trained ch	×	0.08
Models fine-tuned on PRM800K solutions achieve significantly better performance than those fine-tuned on original MATH s	×	0.09
The original solutions in the MATH dataset are more abstract, while the solutions generated by GPT-4 are more fine-grain	×	0.08
The cross-entropy loss following the maximum likelihood estimation (MLE) paradigm is used for fine-tuning.	×	0.07
Two significant gaps in LLMs' math problem solving performance are noted: the gap between greedy-decoding (Pass@1) and m	×	0.13

References

- http://arxiv.org/abs/2510.00071v2

- http://arxiv.org/abs/2509.25160v1
- http://arxiv.org/abs/2310.10047v1