Language Models for Competition-Level Software Engineering Problem Solving

Assignee Research

June 6, 2026

Abstract

This report synthesises findings from 11 peer-reviewed papers addressing the following research question: What techniques enable language models to solve competition-level software engineering problems v15. 12 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: Morescient GAI for Software Engineering (Extended Version). Research question: What techniques enable language models to solve competition-level software engineering problems v15.

2 Methodology

Systematic literature search across multiple databases yielded 11 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

11 papers retrieved. 12 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 benchmark table on page 8 shows the operations and their corresponding results for different data structures.	×	0.02
The operation 'create ' is listed in the first row of the first benchmark table on page 8.	×	0.01
The operation 'TRUE enqueue A1 1' is listed in the second row of the first benchmark table on page 8.	×	0.02
The operation 'TRUE enqueue A1 2' is listed in the third row of the first benchmark table on page 8.	×	0.01
The operation '1 peek A1' is listed in the fourth row of the first benchmark table on page 8.	×	0.01
The operation '2 size A1' is listed in the fifth row of the first benchmark table on page 8.	×	0.01
The operation '1 dequeue A1' is listed in the sixth row of the first benchmark table on page 8.	×	0.01
The operation '1 size A1' is listed in the seventh row of the first benchmark table on page 8.	×	0.01
The operation 'create 2' is listed in the first row of the second benchmark table on page 8.	×	0.02
The operation '3 sum A1 1 2' is listed in the second row of the second benchmark table on page 8.	×	0.02
The benchmark table on page 8 includes columns for 'Time taken', 'Memory used', and 'Branches Covered'.	×	0.03
The third benchmark table on page 8 shows a matrix of values for different data structures and operations.	×	0.01

References

- `http://arxiv.org/abs/1707.03869v3`

- http://arxiv.org/abs/2406.04710v2
- http://arxiv.org/abs/2303.12869v1