Context Window Truncation Strategies and Defect Localization Accuracy in LLaMA 3.2 on BugsInPy

Assignee Research

June 8, 2026

Abstract

This report synthesises findings from 7 peer-reviewed papers addressing the following research question: What is the correlation between context window truncation strategies and defect localization accuracy for locally deployed LLaMA 3.2 models on the BugsInPy dataset. 8 claims were extracted from source literature; 8 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 8.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: A Quantitative and Qualitative Evaluation of LLM-Based Explainable Fault Localization. Research question: What is the correlation between context window truncation strategies and defect localization accuracy for locally deployed LLaMA 3.2 models on the BugsInPy dataset?.

2 Methodology

Systematic literature search across multiple databases yielded 7 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 8.7/10.

3 Results

7 papers retrieved. 8 claims extracted; 8 independently verified. Quality review score: 8.7/10.

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
Existing automated fault localization techniques fail to provide rationales for suggested locations.	✓	0.25
AutoFL is a Large Language Model (LLM)-based fault localization technique that generates an explanation of the bug along	✓	0.33
AutoFL prompts an LLM to use function calls to navigate a repository.	✓	0.28
AutoFL is designed to overcome the limit of LLM context length when localizing faults over large software repositories.	✓	0.19
Experiments were conducted on 798 real-world bugs in Java and Python.	✓	0.17
AutoFL improves method-level acc@1 by up to 233.3% over baselines.	✓	0.24
Developers interviewed generally liked the natural language explanations generated by AutoFL.	✓	0.22
Developers interviewed preferred reading a few high-quality explanations instead of many.	✓	0.24

References

- https://openalex.org/W7159897974
- https://doi.org/10.1109/icst62969.2025.10988982
- https://doi.org/10.48550/arxiv.2308.05487