DeepSeek-R1 and Claude Performance on SWE-Bench Verified With and Without File Context

Assignee Research

May 30, 2026

Abstract

This report synthesises findings from 3 peer-reviewed papers addressing the following research question: What is the performance difference between DeepSeek-R1 and Claude models on SWE-bench Verified when evaluated with and without access to issue-specific file context. Code repair is a fundamental task in software development, facilitating efficient bug resolution and software maintenance. Although large language models (LLMs) have demonstrated considerable potential in automated code repair, their ability to comprehend and leverage diverse. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.0/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: FeedbackEval: A Benchmark for Evaluating Large Language Models in Feedback-Driven Code Repair Tasks. Research question: What is the performance difference between DeepSeek-R1 and Claude models on SWE-bench Verified when evaluated with and without access to issue-specific file context?.

2 Methodology

Systematic literature search across multiple databases yielded 3 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.0/10.

3 Results

3 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 4.0/10.

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

- https://arxiv.org/abs/2508.21433
- https://arxiv.org/abs/2504.06939
- https://arxiv.org/abs/2503.22424