

Training Data Contamination Effects on Qwen3 Model Performance Across Scales on SWE-Bench Verified

Assignee Research

May 30, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: What is the impact of training data contamination on Qwen3-235B's performance across different model sizes on SWE-bench Verified. Abstract The rapid evolution of large language models (LLMs) has driven a transformative shift in artificial intelligence (AI), reshaping both research paradigms and practical applications. Distinguished from their predecessors by unprecedented scale and advanced capabilities. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: A Survey of Large Language Models. Research question: What is the impact of training data contamination on Qwen3-235B's performance across different model sizes on SWE-bench Verified?.

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 6.5/10.

3 Results

15 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 6.5/10.

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://doi.org/10.1007/s11704-026-60308-3>
- <https://doi.org/10.48550/arxiv.2210.11416>
- <https://doi.org/10.48550/arxiv.2505.09388>