

DeepSeek-V3 Scaling and Consistency on Out-of-Distribution Reasoning Benchmarks

Assignee Research

May 30, 2026

Abstract

This report synthesises findings from 8 peer-reviewed papers addressing the following research question: What is the correlation between model scaling and consistency metrics for DeepSeek-V3 when evaluated on out-of-distribution reasoning benchmarks. Recently, there is a high demand for deploying DeepSeek-R1 and V3 locally, possibly because the official service often suffers from being busy and some organizations have data privacy concerns. While single-machine deployment offers infrastructure simplicity, the models' 671B. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 1.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Quantitative Analysis of Performance Drop in DeepSeek Model Quantization. Research question: What is the correlation between model scaling and consistency metrics for DeepSeek-V3 when evaluated on out-of-distribution reasoning benchmarks?.

2 Methodology

Systematic literature search across multiple databases yielded 8 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 1.7/10.

3 Results

8 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 1.7/10.

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.

References

- <http://arxiv.org/abs/2412.07942v1>
- <http://arxiv.org/abs/2402.04177v3>
- <http://arxiv.org/abs/2505.02390v2>