DeepSeek-R1 and Claude-3 Token Efficiency in Few-Shot Code Generation on HumanEval

Assignee Research

May 30, 2026

Abstract

This report synthesises findings from 12 peer-reviewed papers addressing the following research question: How does the token efficiency of DeepSeek-R1 compare to Claude-3 when performing few-shot code generation on HumanEval, measured by pass@1 accuracy per token consumed. How far are Large Language Models (LLMs) in performing deep relational reasoning? In this paper, we evaluate and compare the reasoning capabilities of three cutting-edge LLMs, namely, DeepSeek-R1, DeepSeek-V3 and GPT-4o, through a suite of carefully designed benchmark tasks in. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.3/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Are Large Language Models Capable of Deep Relational Reasoning? Insights from DeepSeek-R1 and Benchmark Comparisons. Research question: How does the token efficiency of DeepSeek-R1 compare to Claude-3 when performing few-shot code generation on HumanEval, measured by pass@1 accuracy per token consumed?.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.3/10.

3 Results

12 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.3/10.

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/2410.12381v3
- http://arxiv.org/abs/2506.23128v1
- http://arxiv.org/abs/2306.08568v2