

Impact of Retrieved Passage Count on RETRO Code Generation Accuracy in HumanEval

Assignee Research

June 1, 2026

Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: What is the impact of varying the number of retrieved passages on RETRO’s performance in code generation tasks measured by HumanEval execution accuracy. Large language models have shown remarkable aptitude in code generation, but still struggle to perform complex tasks. Self-repair – in which the model debugs and repairs its own code – has recently become a popular way to boost performance in these settings. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.3/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Is Self-Repair a Silver Bullet for Code Generation?. Research question: What is the impact of varying the number of retrieved passages on RETRO’s performance in code generation tasks measured by HumanEval execution accuracy?.

2 Methodology

Systematic literature search across multiple databases yielded 14 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.3/10.

3 Results

14 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 5.3/10.

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/2101.00294v3>
- <http://arxiv.org/abs/2402.12317v2>
- <http://arxiv.org/abs/2306.09896v5>