How does 4-bit quantization influence the alignment of LLaMA 3.2 and Mistral with human preferences in code re

Assignee Research

June 10, 2026

Abstract

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. However, despite its increasing popularity, existing studies of self-repair have been limited in scope; in many settings, its efficacy thus remains poorly understood. In this paper, we analyze Code Llama, GPT-3.5 and GPT-4's ability to perform self-repair on problems taken from HumanEval and APPS. We find that when the cost of carrying o

1 Introduction

This paper examines: Is Self-Repair a Silver Bullet for Code Generation?. Research question: How does 4-bit quantization influence the alignment of LLaMA 3.2 and Mistral with human preferences in code repair tasks as evaluated by the TruthfulQA benchmark?.

2 Methodology

Systematic literature search across multiple databases yielded 11 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

11 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

- http://arxiv.org/abs/2402.18571v3
- http://arxiv.org/abs/2310.06825v1
- http://arxiv.org/abs/2306.09896v5