

# Parameter Scale Impact on Solution Consistency in Decoder-Only Models for HumanEval Pro

Assignee Research

June 9, 2026

## Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: To what extent does increasing the parameter scale of decoder-only models improve their ability to maintain solution consistency between base and derived problems in the HumanEval Pro benchmark. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: HumanEval Pro and MBPP Pro: Evaluating Large Language Models on Self-invoking Code Generation. Research question: To what extent does increasing the parameter scale of decoder-only models improve their ability to maintain solution consistency between base and derived problems in the HumanEval Pro benchmark?.

## 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 3.2/10.

## 3 Results

14 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.2/10.

## 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/2402.04177v3>
- <http://arxiv.org/abs/2410.12381v3>
- <http://arxiv.org/abs/2412.21199v2>