Dynamic Reward Scaling vs. Human-Crafted Unit Tests in Code Generation Benchmarks

Assignee Research

June 1, 2026

Abstract

This report synthesises findings from 9 peer-reviewed papers addressing the following research question: How does dynamic reward scaling perform relative to human-crafted unit tests in terms of code correctness and inference latency when evaluated on the HumanEval and SQuTR benchmarks using a fixed. Current large language models (LLMs) often struggle to produce accurate responses on the first attempt for complex reasoning tasks like code generation. Prior research tackles this challenge by generating multiple candidate solutions and validating them with LLM-generated unit. 8 claims were extracted from source literature; 3 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Dynamic Scaling of Unit Tests for Code Reward Modeling. Research question: How does dynamic reward scaling perform relative to human-crafted unit tests in terms of code correctness and inference latency when evaluated on the HumanEval and SQuTR benchmarks using a fixed compute budget?.

2 Methodology

Systematic literature search across multiple databases yielded 9 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.8/10.

3 Results

9 papers retrieved. 8 claims extracted; 3 independently verified. Quality review score: 5.8/10.

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.

5 Extracted Claims

Claim	Verified	Confidence
CodeRM-8B significantly improves the performance of smaller models (e.g., a performance gain of 18.43% on HumanEval Plus	✓	0.18
CodeRM-8B enhances the performance of significantly larger models or even proprietary models (e.g., a 4.95% gain for Lla	×	0.15
Dynamic unit test scaling brings additional performance improvements at a fixed computational cost (e.g., up to approxim	×	0.07
There is a positive correlation between the number of unit tests and the quality of the code reward signal, with greater	✓	0.29
Gemma-2-27B-it performs significantly worse than Llama3.1-70B with a single unit test, but achieves comparable performan	×	0.07
Scaling unit tests is more effective for harder problems.	✓	0.15
The unit test-based majority voting framework follows a standard best-of-N strategy.	×	0.04
The optimal candidate solution Copt is selected based on majority voting, which determines Copt as the solution that pas	×	0.08

References

- http://arxiv.org/abs/2501.01054v1

- http://arxiv.org/abs/2102.07660v2
- http://arxiv.org/abs/2602.17684v2