

# Iterative Self-Refinement vs. Single-Pass Decoding in CodeGen-2B on HumanEval

Assignee Research

June 6, 2026

## Abstract

This report synthesises findings from 11 peer-reviewed papers addressing the following research question: What is the difference in pass@k metrics between iterative self-refinement and single-pass decoding for codegen-2b on the HumanEval benchmark. 12 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: HPL-MxP Benchmark: Mixed-Precision Algorithms, Iterative Refinement, and Scalable Data Generation. Research question: What is the difference in pass@k metrics between iterative self-refinement and single-pass decoding for codegen-2b on the HumanEval 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 3.8/10.

## 3 Results

11 papers retrieved. 12 claims extracted; 0 independently verified. Quality review score: 3.8/10.

## 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
The formula to compute figures of merit for HPL-MxP is $(2/3) * n^3 * t^{-1}$ , where $n$ is the input matrix size and $t$ is the	×	0.15
Monotonic summation, defined as $x < x\_bar$ and $y < y\_bar$ implying $FP(x + y) \leq FP(x\_bar + y\_bar)$ , is requested by the IEE	×	0.01
Monotonic summation may be violated in recent mixed-precision hardware despite being requested by the IEEE 754 standard.	×	0.05
Bisection eigenvalue methods are accurate and numerically stable with only mild assumptions about the monotonicity of $fl$	×	0.01
For very large matrix sizes in mixed-precision contexts, intermediate values in the Sturm function during bisection eige	×	0.05
Iterative refinement was originally deemed beneficial provided the accumulation of residual vector entries is performed	×	0.04
Direct sparse solvers have utilized up to five precisions in refinement.	×	0.02
Cholesky QR allows utilization of a higher precision format to maintain accuracy otherwise not possible in a uniform pre	×	0.02
At the time of the study, mixed-precision algorithms were asymptotically 2x faster than uniform precision algorithms for	×	0.09
In mixed-precision iterative refinement, accuracy can be guaranteed with low values of the condition number of the origi	×	0.11
Ill-conditioning was the original design constraint of the GMRES-IR algorithm.	×	0.03
It is possible to use 3 or even 5 precisions in mixed-precision solver frameworks.	×	0.08

## References

- <http://arxiv.org/abs/2601.15286v1>
- <http://arxiv.org/abs/2410.12381v3>
- <http://arxiv.org/abs/2509.19618v1>