

# INT4 Quantization Effects on Zero-Shot Code Generation in Llama-3.1 Variants

Assignee Research

May 31, 2026

## Abstract

This report synthesises findings from 14 peer-reviewed papers addressing the following research question: How does INT4 quantization impact the zero-shot code generation performance of Llama-3.1-70B compared to smaller variants (e.g., 8B) on HumanEval, and does the trade-off scale with model size. Recent progress in large-scale zero-shot speech synthesis has been significantly advanced by language models and diffusion models. However, the generation process of both methods is slow and computationally intensive. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.3/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: FlashSpeech: Efficient Zero-Shot Speech Synthesis. Research question: How does INT4 quantization impact the zero-shot code generation performance of Llama-3.1-70B compared to smaller variants (e.g., 8B) on HumanEval, and does the trade-off scale with model size?.

## 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 3.3/10.

## 3 Results

14 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.3/10.

## 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/2411.10958v7>
- <http://arxiv.org/abs/2404.14700v4>
- <http://arxiv.org/abs/2410.05265v2>