# Quantization Trade-offs in Sub-10B Parameter Language Models on SLM-Bench

Assignee Research

June 8, 2026

**Abstract**

This report synthesises findings from 9 peer-reviewed papers addressing the following research question: What is the impact of model quantization techniques on the accuracy-throughput trade-off for SLMs under 10B parameters when evaluated on SLM-Bench across different hardware configurations. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: SLM-Bench: A Comprehensive Benchmark of Small Language Models on Environmental Impacts–Extended Version. Research question: What is the impact of model quantization techniques on the accuracy-throughput trade-off for SLMs under 10B parameters when evaluated on SLM-Bench across different hardware configurations?.

## 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 3.7/10.

## 3 Results

9 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.7/10.

## 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/2508.15478v2
- http://arxiv.org/abs/2601.08844v1
- http://arxiv.org/abs/2412.02602v1