

Performance-Efficiency Ratio vs. BLEU and ROUGE in Small Language Model Summarization

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: How does the Performance-Efficiency Ratio (PER) metric compare to traditional evaluation metrics like BLEU or ROUGE when applied to small language models (e.g., 0.5B-1B parameters) on summarization. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Task-Specific Efficiency Analysis: When Small Language Models Outperform Large Language Models. Research question: How does the Performance-Efficiency Ratio (PER) metric compare to traditional evaluation metrics like BLEU or ROUGE when applied to small language models (e.g., 0.5B-1B parameters) on summarization tasks like CNN/DailyMail?.

2 Methodology

Systematic literature search across multiple databases yielded 15 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.8/10.

3 Results

15 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 6.8/10.

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.

References

- <http://arxiv.org/abs/2603.21389v1>
- <http://arxiv.org/abs/2507.08342v1>
- <http://arxiv.org/abs/2408.13006v2>