

Quantization Impact on Token Throughput and Fault Classification in Llama3.1 and Mistral 7B for Battery Management Systems

Assignee Research

June 4, 2026

Abstract

This report synthesises findings from 3 peer-reviewed papers addressing the following research question: What is the effect of model quantization levels on the token throughput and fault classification performance of Llama3.1 and Mistral 7B when applied to battery management system datasets. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: A Review of Large Language Models for Energy Systems: Applications, Challenges, and Future Prospects. Research question: What is the effect of model quantization levels on the token throughput and fault classification performance of Llama3.1 and Mistral 7B when applied to battery management system datasets?.

2 Methodology

Systematic literature search across multiple databases yielded 3 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

3 papers retrieved. 0 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.

References

- <https://doi.org/10.1109/access.2025.3610994>
- <https://openalex.org/W7104183491>
- <https://doi.org/10.48550/arxiv.2502.01703>