Fine-Tuning Effects on Llama3.1 and Mistral 7B Robustness in RAG-Enhanced Power Grid Anomaly Detection

Assignee Research

June 4, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: Comparison of Llama3.1 and Mistral 7B in power grid anomaly detection: How does fine-tuning on battery datasets affect their robustness (F1-score) on downstream tasks when integrated with. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Anomaly Detection: How to Artificially Increase your F1-Score with a Biased Evaluation Protocol. Research question: Comparison of Llama3.1 and Mistral 7B in power grid anomaly detection: How does fine-tuning on battery datasets affect their robustness (F1-score) on downstream tasks when integrated with retrieval-augmented generation (RAG)?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.2/10.

3 Results

13 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 5.2/10.

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/2106.16020v1
- http://arxiv.org/abs/1501.07329v4
- http://arxiv.org/abs/2406.16308v1