

Llama-3-8B-128K Performance on Jamendo-MT-QA Against Open-Source LLMs

Assignee Research

May 30, 2026

Abstract

This report synthesises findings from 10 peer-reviewed papers addressing the following research question: How does the performance of Llama-3-8B-128K compare to other open-source LLMs (e.g., Falcon-40B, Mistral-7B) on Jamendo-MT-QA when evaluated using both human annotations and automated metrics like. Recently, large language models (LLMs) have been deployed to tackle various software engineering (SE) tasks like code generation, significantly advancing the automation of SE tasks. However, assessing the quality of these LLM-generated code and text remains challenging. 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: Can LLMs Replace Human Evaluators? An Empirical Study of LLM-as-a-Judge in Software Engineering. Research question: How does the performance of Llama-3-8B-128K compare to other open-source LLMs (e.g., Falcon-40B, Mistral-7B) on Jamendo-MT-QA when evaluated using both human annotations and automated metrics like ROUGE-L and BLEU?.

2 Methodology

Systematic literature search across multiple databases yielded 10 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

10 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/2502.06193v3>
- <http://arxiv.org/abs/2603.12895v1>
- <http://arxiv.org/abs/2406.03339v2>