Lug-Llama-8B-wura vs. Emu2 and Qwen-VL on GSM8K-V Under Adversarial Visual Perturbations

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 11 peer-reviewed papers addressing the following research question: How does the performance of Lug-Llama-8B-wura compare to Emu2 and Qwen-VL on the GSM8K-V benchmark when evaluated under adversarial visual perturbations, and what is the impact on mathematical. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: GSM8K-V: Can Vision Language Models Solve Grade School Math Word Problems in Visual Contexts. Research question: How does the performance of Lug-Llama-8B-wura compare to Emu2 and Qwen-VL on the GSM8K-V benchmark when evaluated under adversarial visual perturbations, and what is the impact on mathematical reasoning accuracy?.

2 Methodology

Systematic literature search across multiple databases yielded 11 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.5/10.

3 Results

11 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 6.5/10.

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/2509.25160v1
- http://arxiv.org/abs/2308.12966v3
- http://arxiv.org/abs/2511.06522v1