Qwen2-VL Robustness to Visual Noise in GSM8K-V Benchmarking Against LLaVA-1.6 and InternVL

Assignee Research

June 6, 2026

Abstract

This report synthesises findings from 3 peer-reviewed papers addressing the following research question: How does Qwen2-VL's accuracy on GSM8K-V compare to LLaVA-1.6 and InternVL when evaluating robustness to visual noise in grade school math problems. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 0.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: The Use of Peremptory Challenges in Capital Murder Trials: A Legal and Empirical Analysis. Research question: How does Qwen2-VL's accuracy on GSM8K-V compare to LLaVA-1.6 and InternVL when evaluating robustness to visual noise in grade school math problems?.

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 0.2/10.

3 Results

3 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 0.2/10.

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

- https://openalex.org/W750789734
- https://doi.org/10.3390/bioengineering12121330
- https://doi.org/10.48550/arxiv.2411.14522