Gemini 1.5 Pro vs. Prior Multimodal Models in Long-Form Video Question Answering

Assignee Research

June 6, 2026

Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: What is the performance delta between Gemini 1.5 Pro and previous multimodal models on video question-answering tasks involving hour-long inputs. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 6.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: LongVideoBench: A Benchmark for Long-context Interleaved Video-Language Understanding. Research question: What is the performance delta between Gemini 1.5 Pro and previous multimodal models on video question-answering tasks involving hour-long inputs?.

2 Methodology

Systematic literature search across multiple databases yielded 16 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 6.8/10.

3 Results

16 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 6.8/10.

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://arxiv.org/abs/2407.15754
- https://arxiv.org/abs/2603.29252
- http://arxiv.org/abs/2411.04998v1