Retrieval Accuracy Degradation in Gemini Models Beyond 500k-Token Contexts

Assignee Research

June 6, 2026

Abstract

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: How does the retrieval accuracy of Gemini 1.5 Flash degrade on the Needle In A Haystack benchmark compared to Gemini 1.5 Pro when context length exceeds 500k tokens. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 5.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Expected Performance of the ATLAS Experiment - Detector, Trigger and Physics. Research question: How does the retrieval accuracy of Gemini 1.5 Flash degrade on the Needle In A Haystack benchmark compared to Gemini 1.5 Pro when context length exceeds 500k tokens?.

2 Methodology

Systematic literature search across multiple databases yielded 4 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 5.8/10.

3 Results

4 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 5.8/10.

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

- http://arxiv.org/abs/1710.05839v2
- http://arxiv.org/abs/2601.07595v3
- http://arxiv.org/abs/0901.0512v4