

Multimodal Model Hallucination Rates and Context Window Length Under Adversarial Noise

Assignee Research

June 1, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: What is the correlation between context window length and hallucination rates in multimodal models when evaluated on adversarial noise injections within the MM-Needle dataset. Large Language Models (LLMs) have demonstrated exceptional proficiency in instruction-following, becoming increasingly crucial across various applications. However, this capability brings with it the risk of prompt injection attacks, where attackers inject instructions into. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Evaluating the Instruction-Following Robustness of Large Language Models to Prompt Injection. Research question: What is the correlation between context window length and hallucination rates in multimodal models when evaluated on adversarial noise injections within the MMNeedle dataset?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.8/10.

3 Results

13 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 4.8/10.

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/2308.10819v3>
- <http://arxiv.org/abs/2110.03135v4>
- <http://arxiv.org/abs/2405.18770v6>