

Model Size and Hallucination Frequency in Open-Source LLMs Across Religious and Scientific QA Tasks

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: What is the correlation between model size and hallucination frequency in open-source LLMs when evaluated on domain-specific religious question-answering datasets compared to scientific QA tasks. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 2.2/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Reassessing Extractive QA Datasets at Scale: LLM-as-a-Judge and In-Depth Analyses. Research question: What is the correlation between model size and hallucination frequency in open-source LLMs when evaluated on domain-specific religious question-answering datasets compared to scientific QA tasks?.

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 2.2/10.

3 Results

13 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 2.2/10.

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

- <http://arxiv.org/abs/2509.22490v1>
- <http://arxiv.org/abs/2504.11972v3>
- <http://arxiv.org/abs/2604.18404v1>