# Retrieval-Augmented 7B vs. Larger Models in Cross-Domain Summarization Consistency

Assignee Research

June 8, 2026

**Abstract**

This report synthesises findings from 13 peer-reviewed papers addressing the following research question: How do retrieval-augmented 7B models compare to larger models (e.g., 13B, 30B) in terms of factual consistency and hallucination rates when evaluated on cross-domain summarization tasks (e.g.,. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: PlainQAFact: Retrieval-augmented Factual Consistency Evaluation Metric for Biomedical Plain Language Summarization. Research question: How do retrieval-augmented 7B models compare to larger models (e.g., 13B, 30B) in terms of factual consistency and hallucination rates when evaluated on cross-domain summarization tasks (e.g., biomedical vs. legal) using metrics like entailment-based evaluation?.

## 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 3.7/10.

## 3 Results

13 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 3.7/10.

## 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/2503.16581v1
- http://arxiv.org/abs/2504.19565v3
- http://arxiv.org/abs/2503.08890v4