

# Multimodal vs. Text-Only RAG Models in Factual Accuracy on TruthfulQA

Assignee Research

June 9, 2026

## Abstract

This report synthesises findings from 4 peer-reviewed papers addressing the following research question: How do fine-tuned multimodal models compare to text-only retrieval-augmented generation (RAG) models in terms of factual accuracy on the TruthfulQA benchmark when evaluated across diverse scientific. 8 claims were extracted from source literature; 6 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 7.5/10. This report is a machine-generated literature synthesis and does not constitute original research.

## 1 Introduction

This paper examines: LLMs-as-Judges: A Comprehensive Survey on LLM-based Evaluation Methods. Research question: How do fine-tuned multimodal models compare to text-only retrieval-augmented generation (RAG) models in terms of factual accuracy on the TruthfulQA benchmark when evaluated across diverse scientific domains?.

## 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 7.5/10.

## 3 Results

4 papers retrieved. 8 claims extracted; 6 independently verified. Quality review score: 7.5/10.

## 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.

## 5 Extracted Claims

Claim	Verified	Confidence
The paper presents a comprehensive survey of the LLMs-as-judges paradigm covering five key perspectives: Functionality,	✓	0.35
The LLMs-as-judges framework is characterized by excellent effectiveness, ability to generalize across tasks, and interp	✓	0.29
The paper provides a systematic definition of LLMs-as-Judges and introduces their functionality.	×	0.15
The paper addresses methodology to construct an evaluation system with LLMs.	×	0.15
The paper investigates potential domains for the application of LLM judges.	✓	0.17
The paper discusses methods for evaluating LLM judges in various contexts.	✓	0.18
The paper provides a detailed analysis of the limitations of LLM judges and discusses potential future directions.	✓	0.21
A relevant resource list for LLMs-as-Judges is maintained at <a href="https://github.com/CSHaitao/Awesome-LLMs-as-Judges">https://github.com/CSHaitao/Awesome-LLMs-as-Judges</a> .	✓	0.25

## References

- <https://doi.org/10.48550/arxiv.2412.05579>
- <https://doi.org/10.3390/bdcc9120320>
- <https://doi.org/10.20944/preprints202601.1138.v1>