Adaptive Retrieval Mechanisms in Multi-Agent Debate Systems for Hallucination Reduction on MultiFC

Assignee Research

June 11, 2026

Abstract

The increasing diversity and scale of video data demand retrieval systems capable of multimodal understanding, adaptive reasoning, and domain-specific knowledge integration. This paper presents LLandMark, a modular multi-agent framework for landmark-aware multimodal video retrieval to handle real-world complex queries. The framework features specialized agents that collaborate across four stages: query parsing and planning, landmark reasoning, multimodal retrieval, and reranked answer synthesis. A key component, the Landmark Knowledge Agent, detects cultural or spatial landmarks and reformulat

1 Introduction

This paper examines: LLandMark: A Multi-Agent Framework for Landmark-Aware Multimodal Interactive Video Retrieval. Research question: To what extent does the adaptive retrieval mechanism in multi-agent debate systems reduce hallucination rates compared to static retrieval methods when evaluated on the MultiFC dataset?.

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 7.4/10.

3 Results

4 papers retrieved. 24 claims extracted; 19 independently verified. Quality review score: 7.4/10.

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
HCMAIC 2025 provides a 250 GB corpus of broadcast and documentary videos.	×	0.13
The HCMAIC 2025 dataset spans domains including news, education, travel, culture, and sports.	×	0.14
A significant portion of the HCMAIC 2025 dataset consists of news and reportage content.	×	0.13
The challenge defines three evaluation tasks: Textual-KIS, Visual Question Answering (QA), and Temporal Reasoning and Ke	✓	0.31
Textual-KIS involves retrieval based on textual keywords extracted from speech, on-screen text, and semantic description	✓	0.18
Visual QA requires frame-level reasoning to identify visual evidence supporting answers to natural-language questions.	✓	0.17
TRAKE requires the retrieval of temporally ordered video segments for event-sequence detection.	✓	0.15
All tasks adopt a unified submission format where each prediction includes the video name and corresponding frame indice	✓	0.20
The framework implements TransNetV2 to extract keyframes from shots rather than processing all video frames.	×	0.10
The keyframe selection algorithm selects three representative keyframes from each shot based on percentiles [0.15, 0.5,	✓	0.16
System performance is measured using the Mean of Top-k R-Scores, averaged across k $\in${1, 5, 20, 50, 100}.	✓	0.26
For Textual-KIS, a result is correct if the predicted video matches the ground-truth video name and the retrieved frame	✓	0.34
For Visual QA, the textual answer must exactly match the reference answer in addition to frame alignment.	✓	0.19
For Temporal Reasoning, the R-Score is computed as the proportion of predicted frames overlapping with the ground-truth	✓	0.25
LLandMark ranked within the top 56 teams selected for official qualification in HCMAIC 2025.	×	0.15
In Round 1 of the HCMAIC 2025 qualifier, LLandMark achieved a score of 20.00 out of a maximum of 23.00.	✓	0.20
In Round 2 of the HCMAIC 2025 qualifier, LLandMark achieved a score of 28.20 out of a maximum of 30.00.	✓	0.18
In Round 3 of the HCMAIC 2025 qualifier,	✓	0.19

References

- http://arxiv.org/abs/2603.02888v1
- http://arxiv.org/abs/2504.05181v2
- http://arxiv.org/abs/2604.07863v1