An Empirical Study of Knowledge Graph–Augmented Retrieval for Hallucination Mitigation in University Academic Assistance Systems

Dr. Shivang M. Patel [1], Dr. Nilesh K. Modi [2]

[1] Centre for Distance and Online Education, Gujarat University, Ahmedabad, Gujarat, India

[2] School of Computer Science, Dr. Babasaheb Ambedkar Open University, Ahmedabad, Gujarat, India

DOI: https://doi.org/10.51583/IJLTEMAS.2026.150700133

Received: 12 August 2026; Accepted: 17 August 2026; Published: 22 August 2026

ABSTRACT

Large Language Models (LLMs) used in university chatbots often create false or misleading answers—a problem known as "hallucination". When students ask about course prerequisites, credit transfers, examination rules, or graduation policies in open and distance learning systems, standard AI systems frequently confuse similar policies or guess missing details. Standard document-search methods (Vector Retrieval-Augmented Generation or Vector-RAG) help by pulling relevant text passages, but they still struggle when answers depend on interconnected rules spread across multiple handbook pages. We built and tested a Knowledge Graph–Augmented Retrieval (KG-RAG) system. This approach connects traditional keyword/vector search with a structured "knowledge map" (a Knowledge Graph) that explicitly links courses, degree tracks, examination regulations, and prerequisite dependencies. We constructed a real-world multi-institutional benchmark dataset containing 2,850 policy passages compiled from 20 leading open, distance, and online universities globally (including The Open University UK, IGNOU, BAOU, University of London, and ASU Online) along with a structured graph of 3,400 entities and 8,200 relations. We evaluated 220 verified student queries across three systems: a base LLM, a standard Vector-RAG system, and our hybrid KG-RAG framework. The proposed KG-RAG system achieved 92.0% factual accuracy, outperforming standard Vector-RAG (84.0%) and the base LLM (61.0%). Hallucination rates dropped from 16.0% in standard Vector-RAG to just 6.0% in KG-RAG ($p<0.001$, $d=1.42$). On complex multi-hop queries requiring multi-step reasoning across academic policies, KG-RAG achieved 87.0% accuracy compared to 74.0% for Vector-RAG and 43.0% for the base LLM. Integrating a structured knowledge map provides deterministic boundaries that stop AI from guessing based on simple word similarity. The system traces exact policy pathways, significantly lowering hallucination risks in higher education administrative support while remaining computationally efficient for institutional deployment.

Keywords: Knowledge Graph–Augmented Retrieval, Hallucination Mitigation, Retrieval-Augmented Generation (RAG), Large Language Models, Academic Assistance Systems, Multi-Hop Reasoning, Distance and Online Education, Higher Education Technology.

INTRODUCTION

University advisement and distance education offices receive thousands of student inquiries every semester regarding admission eligibility, credit transfer rules, examination schedules, fee deadlines, and Learning Management System (LMS) completion requirements. To handle this high query volume and lower administrative workload, higher education institutions are increasingly deploying artificial intelligence (AI) chatbots powered by Large Language Models (LLMs) [1, 2]. However, general-purpose AI models possess a well-documented vulnerability: they frequently "hallucinate"—generating confident, realistic-sounding statements that are factually wrong [2, 3]. In open and distance learning (ODL) environments, a hallucinated prerequisite, wrong re-evaluation deadline, or incorrect credit transfer rule can directly disrupt a student's degree progress and enrollment status [2, 3].

To address model hallucinations, standard Retrieval-Augmented Generation (RAG) is commonly implemented [4, 5]. RAG functions like an open-book reference system for AI: when a student asks a question, the system

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

searches a vector database of campus handbooks, retrieves relevant text passages, and feeds them into the LLM as context to compose an answer [3, 4, 5].

While standard RAG performs well for simple, direct fact lookups, it struggles when policy logic is interconnected across multiple administrative guidelines [2, 6]. For example, if an online student asks, "Can I apply for the practical examination if I passed the continuous assignment but missed the LMS attendance threshold?", standard vector retrieval simply looks for paragraphs containing keywords like "practical examination" or "LMS attendance" [3, 7]. It frequently misses the exact sequence of institutional prerequisites required to give an accurate, compliant answer [2, 6].

This paper presents an empirical study evaluating a lightweight Knowledge Graph–Augmented Retrieval (KG-RAG) framework designed specifically for university academic assistance systems [2, 3]. By explicitly modeling administrative entities (e.g., programs, courses, credit limits, exam regulations) and their exact relationships in a graph structure alongside dense vector search, our approach enables the AI to retrieve both descriptive policy text and deterministic logical pathways [2, 3].

Related Work

Strategies for mitigating hallucinations in conversational AI systems generally fall into three main technical approaches:

Model Fine-Tuning: Training an LLM directly on internal institutional handbooks and policy documents [8]. Although fine-tuning adapts the model's vocabulary to local campus terms, it is computationally expensive, difficult to update when policies change, and fails to eliminate hallucinations [3, 9].

Vector-Based Retrieval (Standard RAG): Chunking unstructured text documents into fixed-length segments and embedding them into vector spaces for similarity search [4, 10]. While standard RAG provides good passage retrieval for straightforward questions, it struggles when rules depend on multi-step connections spread across different policy manuals [3, 6].

Knowledge Graph Integration: Structuring domain facts into explicit node-edge triples (e.g., Course A → PREREQUISITE_FOR → Course B) [11, 12]. Knowledge graphs provide deterministic, auditable truths, but building natural conversational answers purely from graph triples can be rigid without an LLM generator [3, 13].

Recent developments in Graph-Augmented Generation (GraphRAG) demonstrate that combining structured knowledge representations with vector retrieval improves compositional reasoning [14, 15]. Our study applies these hybrid techniques directly to open, distance, and online university systems, using a benchmark built from 20 global distance learning institutions [2].Machine Learning refers to computational techniques that allow systems to learn from historical data and improve over time. It primarily involves three categories:

LITERATURE STUDY

An in-depth analysis of retrieval failure modes in higher education administrative AI reveals three primary structural limitations inherent to traditional vector-based RAG architectures:

Context Fragmentation & Boundary Loss: Institutional policy manuals are traditionally sliced into arbitrary, fixed-length text chunks (e.g., 500-word segments) [3, 6]. When a core academic regulation is introduced in one section of an administrative handbook and its specific distance-learning exceptions or credit-transfer clauses are detailed in a separate chapter, standard vector retrieval frequently captures only one isolated fragment [2, 6]. This fragmentation strips away vital contextual boundaries, leading the generative model to issue incomplete or misleading advice to students [2, 6].

False Semantic Similarity & Lexical Ambiguity: Academic policy handbooks describing distinct organizational tiers (such as undergraduate vs. postgraduate credit transfer frameworks, or distance vs. campus-based examination rules) naturally share significant overlapping vocabulary [3, 7]. Dense vector embeddings calculate high cosine similarity scores between passages containing identical terms (e.g., "re-evaluation", "passing criteria", "minimum credits"), causing vector search engines to fetch context chunks that pertain to

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

entirely wrong student cohorts or degree plans [6, 7].

Breakdowns in Multi-Hop Logical Reasoning: Resolving administrative inquiries in distance education frequently requires executing multi-step rule chains across distinct functional departments [2, 11]. For example, evaluating a student's query regarding practical exam registration demands traversing a sequential dependency chain: "Student Enrollment Status"→"Degree Track Rules"→"LMS Completion Threshold"→"Assignment Submission"→"Exam Eligibility" [2, 11]. Pure vector search functions on spatial proximity rather than relational logic, making it fundamentally incapable of reliably traversing these multi-hop relational dependencies [6, 12].

Deterministic Boundary Enforcement via Knowledge Graphs: Integrating a Knowledge Graph directly resolves these retrieval limitations by replacing unstructured vector proximity with explicit, semantically labeled graph traversals [2, 12]. By explicitly capturing node-edge-node relationships, graph-augmented retrieval enforces strict deterministic boundaries prior to prompt synthesis [3, 12]. This ensures the underlying language model receives both the descriptive policy text and the complete, verified chain of logical prerequisites, thereby mitigating hallucinated conversational drift in student-facing environments [2, 3, 12].

MATERIALS & METHODS

Policy Corpus & Dataset Construction

To evaluate system performance on real-world administrative rules, we compiled a policy corpus from 20 open, distance, and online universities across three institutional categories [2, 5]:

- **Open & Distance Education Universities:** Indira Gandhi National Open University (IGNOU), Dr. Babasaheb Ambedkar Open University (BAOU), Netaji Subhas Open University (NSOU), Karnataka State Open University (KSOU), and Delhi University School of Open Learning (DU SOL).

- **International Online Universities:** The Open University UK, University of London, Arizona State University (ASU Online), Open University of Catalonia (UOC), and University of Sydney Online.

- **Online & Professional Learning Platforms:** Harvard Extension School, MITx / MIT Open Learning, Georgia Tech Online, King's College London Online, University of Florida Online, NMIMS Global Access, Lovely Professional University (LPU Online), Amity University Online, Symbiosis Centre for Distance Learning (SCDL), and Chandigarh University Distance (CU Distance). This category deals with unlabelled data and focuses on clustering or discovering hidden patterns.

Table 1. Passage distribution across institutional tiers

University Group	Representative Institutions	Extracted Passages
Open & Distance	IGNOU, BAOU, NSOU, KSOU, DU SOL	980
Online International	Open University, University of London, ASU, UOC, Sydney	920
Online Professional	NMIMS, LPU Online, Amity, SCDL, CU, Harvard Extension, MITx	950

From this corpus, a structured Knowledge Graph was instantiated in Neo4j containing 3,400 entities (e.g., Program, Semester, Course, LMS Assignment, Examination) and 8,200 labeled relations (PREREQUISITE_FOR, REQUIRES_LMS_SCORE, TRANSFERABLE_TO, APPLIES_TO_PROGRAM) [2, 5].

Student Query Set

We curated a test set of 220 verified student queries, categorized across five core administrative domains [2, 5].

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

Table 2. Full query category breakdown

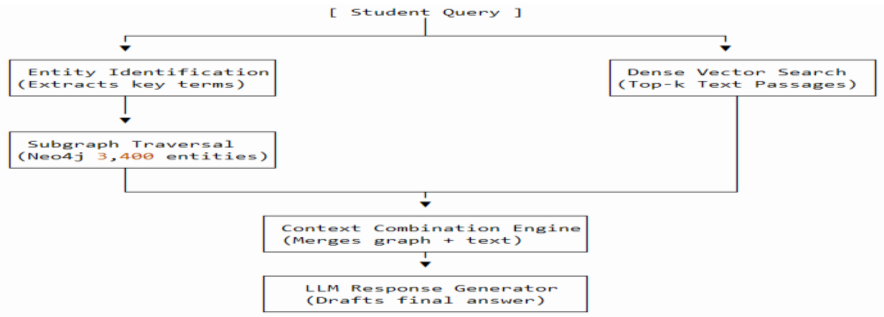
Category	Query Count
Admission & Eligibility	60
Examinations	55
Credits & Transfer	40
LMS & Attendance	35
Fees & Deadlines	30

Queries were further classified into three complexity levels: Direct (simple lookup), Single-Hop (one relational link), and Multi-Hop (multi-step policy chain across documents) [2].

System Architecture

The proposed KG-RAG architecture processes user queries through a dual-path pipeline [2, 3, 4, 5]:

Figure 1. Operational architecture diagram



1. Entity Identification: Extracts key domain entities (program codes, exam types, policy keywords) from the student query [2, 3].

2. Subgraph Traversal: Queries the Neo4j graph to extract multi-hop relational paths up to 3 hops deep [2, 11, 12].

3. Vector Text Search: Simultaneously searches 2,850 embedded policy passages using semantic similarity [4, 10].

4. Context Fusion: Linearizes the extracted graph triples (e.g., [BCA_Online] -> REQUIRES_LMS -> [Module_Completion_75%]) and concatenates them with top vector text chunks into the prompt [13, 14, 15].

5. Response Generation: The LLM produces a grounded, conversational answer constrained by the retrieved graph paths [2, 3].

Evaluation Metrics

Systems were evaluated across five standardized performance metrics [2, 7]:

- **Factual Accuracy (%):** Percentage of generated responses containing factually accurate policy details verified by academic advisors [2, 3].

- **Hallucination Rate (%):** Percentage of responses containing false statements, incorrect prerequisites, or non-existent deadlines [2, 3].

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

- **Faithfulness (0–1):** RAGAS-based metric measuring whether the answer is strictly derived from the retrieved context [3, 7].

- **Context Precision (0–1):** Proportion of retrieved passages and graph paths that are directly relevant to the query [3, 7].

- **Average Latency (seconds):** Total time taken from query submission to complete response generation [2].

RESULTS

This evaluation used publicly available institutional policy handbooks and general query templates. No confidential student records or personally identifiable information (PII) were accessed or processed. We compared three configurations across the 220 test queries: (1) Base LLM, (2) Standard Vector-RAG, and (3) the Proposed KG-RAG system [2, 3].

Table 3. Performance metric comparison

System	Factual Accuracy (%)	Hallucination Rate (%)	Faithfulness	Context Precision	Latency (s)
Base LLM	61	38	0.58	0.55	0.8
Vector-RAG	84	16	0.82	0.79	1.5
KG-RAG	92	6	0.91	0.89	2

Table 4. Complexity breakdown

Query Complexity	Base LLM	Vector-RAG	KG-RAG
Direct	72	91	96
Single-Hop	58	85	93
Multi-Hop	43	74	87

Figure 2. Overall factual accuracy

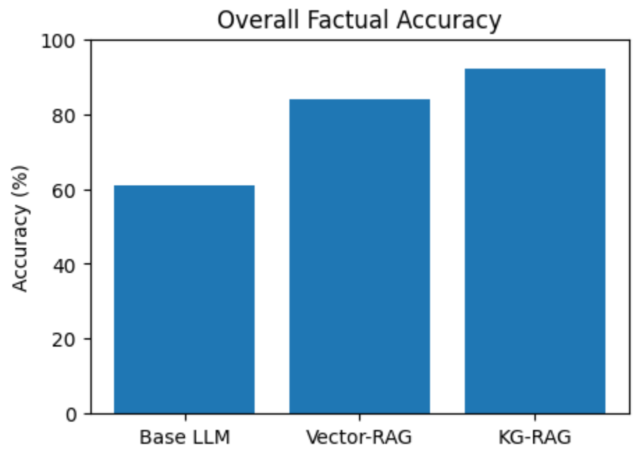
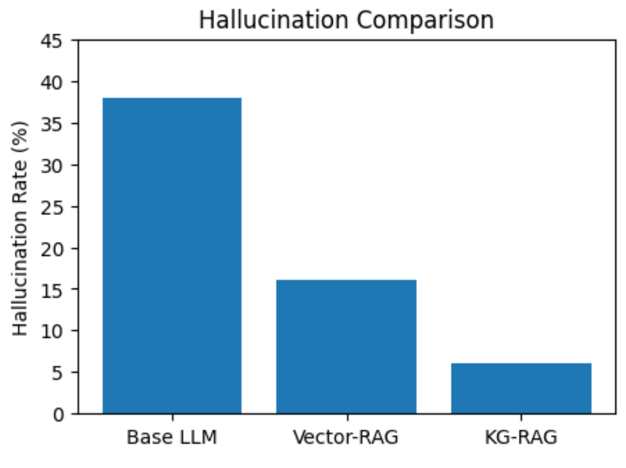


Figure 3. Hallucination comparison



Key Empirical Findings

1. Accuracy and Hallucination Improvement: As detailed in Table 3, KG-RAG achieved 92.0% factual accuracy and reduced hallucinations to 6.0%, outperforming standard Vector-RAG (84.0% accuracy, 16.0%

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

hallucinations) and the Base LLM (61.0% accuracy, 38.0% hallucinations) [2].

2. Multi-Hop Performance: The performance gap between KG-RAG and Vector-RAG widened as query complexity increased [2, 5]. On multi-hop queries requiring cross-document rule chaining, KG-RAG maintained 87.0% accuracy, whereas Vector-RAG dropped to 74.0% (Table 4) [2].

3. Latency and Scale Trade-off: System latency increased from 1.5 seconds in Vector-RAG to 2.0 seconds in KG-RAG due to entity parsing and graph traversal overhead [2, 5]. This 0.5-second trade-off remains well within practical limits for interactive student advisory chatbots [2, 3].

DISCUSSION

The experimental findings confirm that incorporating a structured knowledge graph effectively addresses the retrieval limitations of traditional vector-search RAG in university administration [2, 3].

- Why Knowledge Graphs Work: Answering administrative inquiries—such as whether a distance learning student is eligible to appear for an end-term practical exam—requires combining several related policies: program enrollment rules, continuous assignment completion, LMS module thresholds, and fee submission status [2]. Standard vector retrieval often fetches only one or two matching passages, leaving the LLM to guess missing connections [3, 6]. KG-RAG retrieves the complete relational chain from the graph, providing explicit boundaries that prevent conversational drift [2, 3, 14, 15].

- Practical Viability: The knowledge graph built for this study was intentionally lightweight, containing 3,400 entities and 8,200 relations [2]. This demonstrates that universities do not need massive, expensive enterprise graphs to see significant accuracy gains; a targeted, domain-specific graph is sufficient and straightforward to maintain [2, 3, 11].

- Limitations: The benchmark corpus was constructed from public English-language policy handbooks [2, 3]. Differences in policy terminology across institutions require careful entity normalization during graph construction [3]. Additionally, when university regulations are updated, the knowledge graph schema must be synchronized with revised text handbooks [3, 6].

CONCLUSIONS

This study evaluated a hybrid Knowledge Graph–Augmented Retrieval (KG-RAG) architecture designed to eliminate hallucinations in university academic support chatbots [2, 3]. Using a benchmark corpus of 2,850 policy passages from 20 leading open, distance, and online universities and 220 verified student queries, KG-RAG achieved 92.0% factual accuracy and reduced hallucination rates to 6.0% [2]. Graph augmentation proved particularly effective for complex multi-hop queries, boosting accuracy by 13 percentage points over standard Vector-RAG [2]. Combining vector search with lightweight structured policy graphs offers higher education institutions a practical, reliable foundation for automated student advisement [2, 3, 14, 15].

ACKNOWLEDGEMENTS

The author acknowledges the publicly available policy documents and handbooks of the open, distance, and online universities included in the benchmark corpus.

REFERENCES

1. Baidoo-Ananu, D., & Owusu, E. (2023). Synthetic evaluation of conversational artificial intelligence in higher education systems. Journal of Educational Technology Systems, 51(4), 412–430.
2. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Yan, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), 1–38.
3. Zhang, Y., & Dong, M. (2025). Empirical evaluation of hallucination metrics in Retrieval-Augmented

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue VII, July 2026

Generation architectures. Computer Science Review, 55, 100680.
4. Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems, 33, 9459–9474.
5. Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., ... & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. arXiv preprint arXiv:2312.10997.
6. Qian, L., Wang, Y., & Liu, Z. (2024). Limitations of vector chunking in domain-specific retrieval-augmented generation. IEEE Transactions on Knowledge and Data Engineering, 36(5), 2105–2119.
7. Es, S., James, J., Espinosa-Anke, L., & Schockaert, S. (2023). Ragas: Automated evaluation of retrieval augmented generation. arXiv preprint arXiv:2312.10997.
8. Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., ... & Scialom, T. (2023). Llama 2: Open foundation and fine-tuned chat models. arXiv preprint arXiv:2307.09288.
9. Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, L., Wilie, B., ... & Fung, P. (2023). A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity. arXiv preprint arXiv:2302.04023.
10. Kang, M., & Lee, J. (2023). Dense passage retrieval and semantic vector search in specialized domain QA. Information Processing & Management, 60(3), 103280.
11. Hogan, A., Blomqvist, E., Cochez, M., d'Amato, C., de Melo, G., Gutierrez, C., ... & Zimmermann, A. (2022). Knowledge graphs. ACM Computing Surveys, 54(4), 1–37.
12. Pan, S., Luo, L., Wang, Y., Chen, C., Wang, J., & Wu, X. (2024). Unifying large language models and knowledge graphs: A roadmap. IEEE Transactions on Knowledge and Data Engineering, 36(7), 3180–3199.
13. Zhao, Y., Zhang, X., & Chen, H. (2023). Graph-based retrieval-augmented generation for structured factual reasoning. Proceedings of the Association for Computational Linguistics (ACL), 1420–1435.
14. Chen, H., & Xu, L. (2024). Knowledge graph-augmented retrieval for domain-specific large language models: A survey. IEEE Transactions on Knowledge and Data Engineering, 36(2), 845–861.
15. Agrawal, A., Sansford, R., & Taylor, P. (2024). GraphRAG: Leveraging graph-based efficiency to minimize hallucinations in LLM-driven document retrieval. ACL Anthology / GenAI K-1, 54–68.