

Review

Retrieval-Augmented Generation for Curated Thematic Corpora: A Critical Survey, Bibliometric Evidence, and the ThemePath-RAG Framework

Winda Monika ^{1,*} , Deshinta Arrova Dewi ² , Arbi Haza Nasution ³ , Aytuğ Onan ⁴  and Yohei Murakami ⁵ 

¹ Department of Library Information, Universitas Lancang Kuning, Pekanbaru 28266, Indonesia

² Faculty of Data Science and Information Technology, INTI International University, Bandar Baru Nilai 71800, Malaysia; deshinta.ad@newinti.edu.my

³ Department of Informatics Engineering, Universitas Islam Riau, Pekanbaru 28284, Indonesia; arbi@eng.uir.ac.id

⁴ Department of Computer Engineering, Faculty of Engineering, Izmir Institute of Technology, Izmir 35430, Turkey; aytugonan@iyte.edu.tr

⁵ Faculty of Information Science and Engineering, Ritsumeikan University, Ibaraki 567-8570, Japan; yohei@fc.ritsumei.ac.jp

* Correspondence: windamonika@unilak.ac.id

Abstract

Retrieval-Augmented Generation (RAG) grounds large language models in external evidence, but many RAG systems represent knowledge either as flat text chunks or as automatically constructed indexing graphs. This assumption is incomplete for curated thematic corpora, including religious scriptures, legal codes, clinical guidelines, educational taxonomies, policy documents, and library classification systems, where domain experts have already organized knowledge into thematic paths and citeable canonical units. This paper investigates how RAG can exploit such expert-authored structures while pruning evidence to a compact and query-specific set. We conduct a critical survey supported by a bibliometric analysis of 2815 Scopus-indexed RAG-related records exported on 26 May 2026, of which 2809 records were retained after duplicate removal. The bibliometric results indicate rapid growth in RAG research but limited explicit consolidation around curated thematic paths, canonical evidence units, or thematic path-guided evidence pruning. We therefore propose ThemePath-RAG, a retrieval framework that retrieves curated thematic paths as high-recall semantic routes, expands candidate canonical evidence, and applies query-aware scoring and global pruning before generation. To assess operational feasibility, we implement ThemePath-RAG for Qur’anic question answering and compare it with a Vector RAG baseline on 150 paired questions using RAGAS context relevance with gpt-4o-mini as the LLM evaluator. Both methods return approximately three final ayat per question. Vector RAG achieves higher mean context relevance than ThemePath-RAG (0.920 versus 0.798; $p < 0.001$). Thus, the proof of concept establishes the feasibility of thematic-path-guided retrieval and identifies evidence-selection challenges, rather than demonstrating superiority over conventional vector retrieval. The paper clarifies the framework’s relationship to GraphRAG, LightRAG, HippoRAG, PathRAG, ontology-based RAG, and AI-augmented bibliometric systems, and outlines a language-matched, multi-baseline evaluation agenda for future cross-domain validation.

Keywords: Retrieval-Augmented Generation; large language models; critical survey; curated thematic corpora; evidence pruning; grounded question answering; education quality



Academic Editors: Marina Bagic Babac and Andrea Giovanni Nuzzolese

Received: 1 June 2026

Revised: 24 June 2026

Accepted: 27 June 2026

Published: 7 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Large language models (LLMs) have demonstrated strong performance in natural language understanding, reasoning, dialogue, summarization, and question answering. Nevertheless, their reliance on parametric memory creates persistent limitations, including hallucination, outdated knowledge, weak provenance, and difficulty adapting to domain-specific evidence. Retrieval-Augmented Generation (RAG) addresses these problems by retrieving relevant external knowledge at inference time and conditioning the LLM on retrieved context [1]. As a result, RAG has become a practical and scalable approach for factual grounding, knowledge updates, and domain adaptation.

The recent RAG literature has grown rapidly. Based on a Scopus BibTeX export containing 2815 RAG-related records collected in May 2026, the field expanded sharply from 2023 to 2026. The corpus contains only a small number of records in 2023, but it rapidly increases in 2024 and 2025, with 2026 representing only a partial publication year at the time of data collection. Keyword-assisted screening of titles, abstracts, and keywords indicates that dominant research clusters include evaluation and hallucination mitigation, agentic and multi-hop RAG, graph or knowledge-graph RAG, education, software engineering, medical and biomedical applications, multimodal RAG, legal RAG, and security or privacy-oriented RAG. However, only a small fraction of papers explicitly address religious or sacred text RAG, and even fewer focus on curated thematic structures as a first-class retrieval object.

Most existing RAG methods adopt one of two assumptions. The first assumption is that knowledge should be segmented into flat text chunks, embedded into a vector space, and retrieved through similarity search [2–4]. This assumption underlies many production RAG systems because it is simple, scalable, and domain-agnostic. The second assumption is that, when flat chunks are insufficient, the system should automatically construct an indexing graph from unstructured documents using entity extraction, relation extraction, summarization, or community detection [5–7]. Graph-based RAG methods follow this direction by organizing text into graph structures and retrieving nodes, edges, subgraphs, communities, or paths.

Both assumptions are useful, but incomplete. Many high-value corpora are not merely unstructured text. Religious scriptures, legal codes, clinical guidelines, educational curricula, policy documents, and library taxonomies are often already organized by human experts into curated themes, subthemes, canonical units, commentary layers, and cross-references. For example, a Qur’anic thematic corpus may organize content through paths such as Story → Prophets → Miracle of Prophets → Prophet Muhammad → Isra’ Mi’raj. Such paths are not arbitrary graph edges extracted by a model. They are domain-authored semantic routes. They encode expert interpretation, hierarchical organization, and retrieval intent. However, current RAG systems usually either flatten such corpora into chunks or reconstruct a new graph from text, thereby underutilizing the curated thematic structure that already exists. Indonesian legal codes provide another concrete example of this problem. Recent work on Indonesian legal question answering has modeled the KUH Perdata as a graph-structured corpus over books, chapters, sections, and articles, showing that GraphRAG can improve faithfulness, answer relevancy, and contextual grounding in legal QA [8]. However, this also exposes a broader retrieval challenge: once a relevant legal hierarchy or thematic node is found, the system must still select the specific article, paragraph, or clause that directly answers the query. This motivates the need for query-aware evidence pruning over curated thematic or hierarchical paths.

This limitation becomes more evident in question answering. Suppose a user asks about the evidence for Isra’ Mi’raj in the Qur’an. A thematic retriever may correctly identify the endpoint node Isra’ Mi’raj, but if the system retrieves all verses connected to that node,

it may return tens of candidate verses, many of which are thematically related but not directly responsive to the user question. The problem is no longer lack of recall. The problem is insufficient precision within a retrieved thematic space. In other words, the system retrieves the right theme but too much evidence.

This problem aligns with a broader critique of graph-based RAG. PathRAG argues that many graph-based RAG methods suffer from redundancy in retrieved information rather than insufficiency. It notes that GraphRAG may use all information within certain communities, while LightRAG may use immediate neighbors of query-related nodes, both of which can introduce noise and increase token consumption [9]. PathRAG responds by retrieving key relational paths and converting them into textual prompts. However, PathRAG still assumes an indexing graph constructed from a text database, not a curated thematic hierarchy authored by domain experts.

Similarly, HippoRAG improves retrieval by using LLM-generated knowledge graphs and Personalized PageRank (PPR) to mimic associative long-term memory. It addresses the weakness of conventional RAG, where passages are encoded in isolation and multi-hop associations are difficult to retrieve [10,11]. HippoRAG 2 further improves this direction by integrating passage nodes, query-to-triple retrieval, and recognition-memory filtering to improve contextualization and reduce irrelevant triples [12]. These methods are valuable, but their primary focus is on constructing and activating knowledge graphs from text. They do not specifically formulate the problem of exploiting pre-existing curated thematic paths.

The reference set was substantially expanded beyond foundational RAG papers. In addition to canonical works on dense retrieval, BM25, graph-based retrieval, and memory-inspired RAG, this survey cites more than one hundred Scopus-indexed RAG studies from the 2809 cleaned records to cover review papers, retrieval architectures, chunking strategies, graph and knowledge-graph RAG, agentic RAG, evaluation and hallucination mitigation, and major application domains. This broader citation base allows the paper to function not only as a conceptual position paper but also as a literature-grounded critical survey.

This paper therefore asks the following question: How should RAG systems retrieve and prune evidence when the corpus already contains curated thematic paths and canonical evidence units? To answer this question, we present a critical survey of the RAG literature and propose a new research direction called ThemePath-RAG. ThemePath-RAG treats thematic paths as high-recall semantic routes and performs query-aware evidence pruning over canonical units before generation. Instead of retrieving all graph neighbors, all community content, or all evidence under a theme, ThemePath-RAG retrieves a thematic path first, expands candidate evidence from the path, and prunes the evidence according to the user query. This question is motivated by recent domain-specific RAG studies in Qur'anic studies and Indonesian legal QA, which demonstrate the importance of retrieval grounding in sensitive and structurally organized corpora but do not yet formulate query-aware evidence pruning over curated thematic paths as a general framework [8,13].

This paper is not intended as a general descriptive survey of all RAG techniques. Instead, it presents a critical survey focused on the structural assumptions behind current RAG systems. While prior surveys have categorized RAG by retrieval strategy, generation mechanism, evaluation method, and application domain, this paper examines how existing methods represent knowledge before retrieval. Through this lens, we identify an overlooked class of corpora, namely curated thematic corpora, whose expert-authored thematic paths are not fully exploited by current flat or graph-based RAG methods. Based on this analysis, we propose ThemePath-RAG as a framework-oriented research agenda for query-aware evidence pruning over curated thematic paths.

The contributions of this paper are as follows:

- We provide a critical scoping survey of RAG research supported by a bibliometric analysis of 2815 Scopus-indexed records.
- We analyze the dominant bibliometric signals in the RAG literature, including document type composition, keyword distribution, thematic evolution, and country–keyword relationships.
- We introduce the concept of curated thematic corpora, where knowledge is already organized through human-authored thematic structures.
- We identify a limitation in current flat and graph-based RAG methods: they do not fully exploit curated thematic paths and often retrieve redundant evidence.
- We propose ThemePath-RAG, a query-aware evidence pruning framework for grounded question answering over pre-structured thematic corpora.
- We position Qur’anic question answering as a primary case study and discuss generalization to legal, medical, educational, and library knowledge systems.

2. Methodology

2.1. Research Design

This study adopts a bibliometric-assisted critical survey design. The bibliometric component maps the development of Retrieval-Augmented Generation (RAG) research using Scopus-indexed metadata, including publication type, keyword distribution, thematic evolution, country–keyword relationships, and broad topic clusters. The critical survey component interprets the dominant assumptions of existing RAG methods regarding knowledge representation before retrieval. These two components are then used to motivate the proposed ThemePath-RAG framework. The study is not designed as a clinical systematic review or meta-analysis. Instead, PRISMA 2020 is used as a reporting guide to improve transparency in the identification, duplicate removal, screening, and inclusion of records for bibliometric analysis and critical synthesis [14]. This distinction is important because the bibliometric corpus is used for quantitative metadata mapping, whereas a smaller representative subset of studies is used for qualitative comparison and framework development.

2.2. Data Source and Search Strategy

Scopus was selected as the bibliographic data source because it is a large abstract and citation database that provides structured metadata suitable for bibliometric analysis, including document type, source title, publication year, author keywords, indexed keywords, author affiliation, country, DOI, abstract, and citation information [15,16]. Scopus was also appropriate for this study because the objective was to map the indexed scholarly literature on RAG rather than to perform a full systematic search of all possible preprints, technical reports, and grey literature. The search was conducted on 26 May 2026 using the Scopus title, abstract, and keyword fields. The search was limited to records published between 2023 and 2026 because RAG became a rapidly expanding research topic after the broader adoption of LLM-based retrieval and generation pipelines. The year 2026 was treated as a partial publication year because the search was conducted in May 2026. The Scopus search query was formulated as follows:

TITLE-ABS-KEY(“retrieval augmented generation” OR “retrieval-augmented generation” OR “RAG”) AND PUBYEAR > 2022 AND PUBYEAR < 2027

The metadata were exported from Scopus in BibTeX format. The initial export contained 2815 records. The exported metadata were processed using PyBibX version 5.9.5, a Python library for bibliometric and scientometric analysis that supports raw bibliographic data from Scopus, Web of Science, and PubMed [17]. PyBibX was used for duplicate detection, metadata extraction, and preparation of the bibliometric dataset.

2.3. Eligibility Criteria

The eligibility criteria were defined to ensure that the bibliometric corpus represented RAG-related scholarly research while preserving broad coverage of the field. Records were included in the bibliometric corpus if they met the following criteria:

1. The record was indexed in Scopus.
2. The record was published between 2023 and 2026.
3. The title, abstract, author keywords, or indexed keywords indicated relevance to retrieval-augmented generation, retrieval-enhanced language models, graph-based RAG, knowledge-graph RAG, sparse or hybrid retrieval for LLMs, domain-specific RAG, RAG evaluation, or hallucination mitigation in RAG systems.
4. The record contained sufficient bibliographic metadata for analysis, including title, publication year, document type, source title, and either abstract or keyword information.
5. The record belonged to scholarly document categories such as journal articles, conference papers, or review papers.

Records were not excluded from the bibliometric corpus merely because they were not discussed in the qualitative synthesis. This distinction is important because bibliometric analysis benefits from broad metadata coverage, while critical synthesis requires a smaller set of representative studies. Therefore, records that met the general inclusion criteria remained part of the bibliometric corpus even if they were not selected for detailed qualitative discussion. Records were excluded from detailed critical synthesis if they did not represent a distinct retrieval architecture, evaluation concern, knowledge-representation strategy, or domain-specific RAG application. This means that exclusion from the qualitative synthesis does not mean exclusion from the bibliometric analysis.

2.4. PRISMA-Informed Workflow

The record-selection process followed a PRISMA 2020-informed workflow [14]. Figure 1 summarizes the identification, duplicate removal, screening, and inclusion process. The initial Scopus search returned 2815 records. Duplicate removal was performed using PyBibX [17], resulting in the removal of 6 duplicate records. After duplicate removal, 2809 cleaned records were retained. All 2809 cleaned records were included in the bibliometric analysis. These records were used to analyze document type composition, keyword frequencies, thematic evolution, country–keyword relationships, and descriptive metadata patterns. A second screening step was then conducted to select representative studies for detailed critical synthesis. Screening was based on titles, abstracts, author keywords, indexed keywords, and relevance to the structural assumptions of RAG systems. Priority was given to studies representing the following streams: review-oriented RAG studies, flat or dense retrieval, sparse and hybrid retrieval, chunking and structure-aware preprocessing, graph-based RAG, knowledge-graph RAG, ontology-based RAG, path-based RAG, agentic and multi-hop RAG, evaluation and hallucination mitigation, medical RAG, legal RAG, educational RAG, software engineering RAG, multimodal RAG, security and privacy-oriented RAG, and religious or canonical-text RAG. From the 2809 cleaned records, 143 representative studies were selected for detailed critical synthesis and framework comparison. The remaining 2666 records were not excluded from the study; rather, they remained part of the bibliometric corpus but were not discussed individually in the qualitative synthesis.

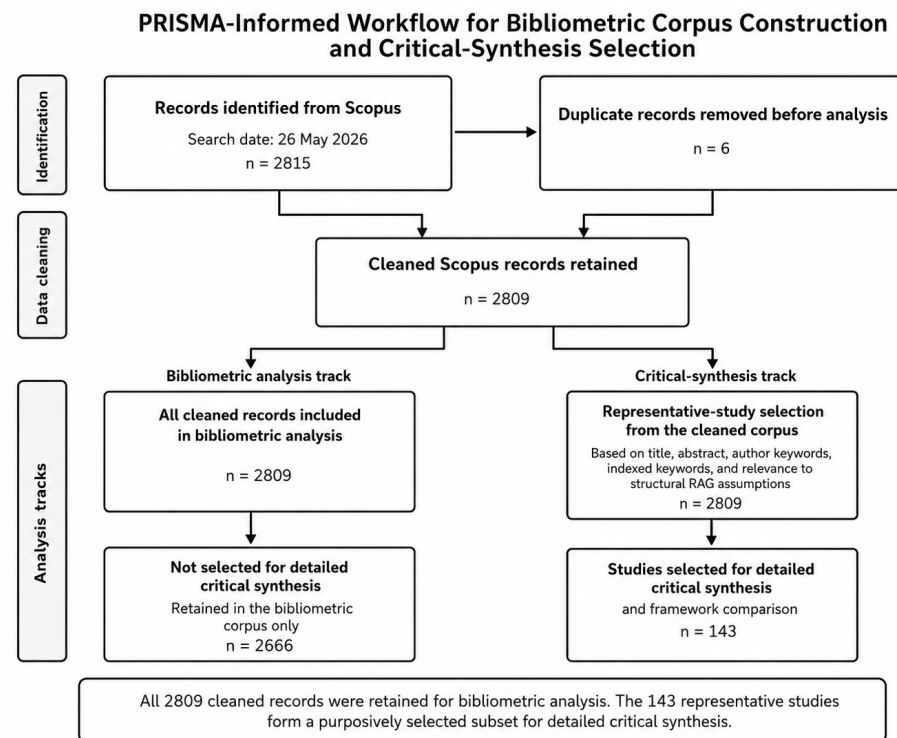


Figure 1. PRISMA-informed workflow for bibliometric corpus construction and representative-study selection for critical synthesis. A total of 2815 Scopus-indexed records were identified, and six duplicate records were removed, resulting in 2809 cleaned records. All cleaned records were retained for bibliometric analysis. From the same corpus, 143 representative studies were purposively selected—based on titles, abstracts, author and indexed keywords, and relevance to the structural RAG assumptions—for detailed critical synthesis and framework comparison.

2.5. Bibliometric Analysis Procedure

The bibliometric analysis was conducted on the 2809 cleaned Scopus records. The analysis focused on descriptive and thematic indicators rather than citation-network analysis because the exported metadata did not include complete reference lists. Consequently, co-citation analysis, bibliographic coupling, and reference publication year spectroscopy were not interpreted as core evidence. Instead, the bibliometric analysis focused on metadata fields that were consistently available across the exported records. The bibliometric analysis included the following components:

1. **Dataset and metadata quality analysis:** This analysis summarized the number of records, publication years, countries, institutions, sources, authors, document types, total citations, average citations per document, and metadata completeness.
2. **Document type analysis:** This analysis examined whether the field was dominated by conference papers, journal articles, or review papers.
3. **Keyword analysis:** Author keywords, indexed keywords, and Keywords Plus were analyzed to identify the dominant research vocabulary in RAG studies.
4. **Thematic evolution analysis:** This analysis examined changes in dominant topics from 2023 to 2026.
5. **Country–keyword relationship analysis:** This analysis examined whether major RAG topics were concentrated in particular countries or shared across leading research communities.
6. **Bibliometric gap synthesis:** The bibliometric results were interpreted to assess whether curated thematic paths, canonical evidence units, and thematic path-guided evidence pruning appeared as consolidated research themes.

The bibliometric results were not treated as standalone proof that existing RAG systems cannot process curated thematic corpora. Instead, they were used as empirical motivation for the critical survey. In particular, the absence of dominant keywords related to curated thematic paths, canonical evidence units, and thematic evidence pruning was interpreted cautiously as evidence of limited explicit consolidation in the Scopus-indexed RAG literature, not as proof of technical impossibility.

2.6. Critical Synthesis Procedure

The critical synthesis was conducted on the 143 representative studies selected from the cleaned bibliometric corpus. These studies were grouped according to their retrieval architecture, knowledge-representation assumptions, and relevance to the ThemePath-RAG research problem. The synthesis was guided by the following question:

How do existing RAG methods represent knowledge before retrieval, and what limitations emerge when the corpus already contains expert-authored thematic paths and canonical evidence units?

Based on this question, the selected studies were organized into several categories: flat chunk-based RAG, sparse and hybrid RAG, graph-based RAG, agentic and multi-hop RAG, memory-inspired and path-based RAG, ontology-based RAG, and domain-specific RAG. The comparison was then used to identify the structural gap addressed by ThemePath-RAG. This synthesis supports the central argument of the paper: current RAG research has advanced significantly in retrieval architecture, graph construction, multi-hop reasoning, and domain adaptation, but the specific setting of RAG over curated thematic corpora remains underdeveloped. ThemePath-RAG is proposed as a conceptual framework for this setting.

2.7. Methodological Limitations

Several methodological limitations should be acknowledged. First, the bibliometric corpus was limited to Scopus-indexed records. Relevant preprints, technical reports, arXiv papers, non-indexed conference papers, and domain-specific repositories may therefore be underrepresented. Second, the 2026 data represent only a partial publication year because the search was conducted in May 2026. Third, bibliometric keyword analysis depends on the terminology used by authors and indexers. The absence of a keyword does not necessarily mean that a concept is technically absent from the literature. Fourth, the representative critical synthesis is selective rather than exhaustive; it was designed to cover major RAG paradigms and application domains rather than to summarize every record in the corpus. Despite these limitations, the PRISMA 2020-informed workflow, explicit eligibility criteria, duplicate-removal procedure, and separation between bibliometric inclusion and qualitative synthesis improve the transparency and reproducibility of the study.

In addition to the bibliometric-assisted critical survey, this study includes an initial Qur'anic QA proof of concept to operationalize the proposed ThemePath-RAG framework. The implementation details, including thematic-path retrieval, weighted evidence scoring, and global evidence pruning, are described in Section 7. The paired evaluation protocol and empirical results are presented in Section 8.

3. Bibliometric Evidence of RAG Research Trends

3.1. Dataset and Metadata Quality

The bibliometric analysis used a Scopus BibTeX export collected on 26 May 2026. As shown in Table 1, the initial export contained 2815 records. After duplicate removal using PyBibX [17], 2809 documents were retained. The dataset spans 2023–2026 and includes 102 countries, 4241 institutions, 1267 sources, and 11,532 authors. The retained records

consist of 1893 conference papers, 877 journal articles, and 39 review papers. The total citation count reported in the dataset is 6816, with an average of 2.43 citations per document.

The dataset has high metadata completeness for sources, affiliations, publication year, authors, author keywords, abstracts, and DOI. However, the reference field was unavailable in the export. Consequently, this paper does not interpret co-citation, bibliographic coupling, or reference publication year spectroscopy as core evidence. The bibliometric results are instead used to support descriptive and thematic mapping, including document type composition, keyword distributions, thematic evolution, country–keyword relationships, and broad topic identification.

Table 1. Bibliometric dataset summary.

Indicator	Value
Initial Scopus records	2815
Records after duplicate removal	2809
Timespan	2023–2026
Countries	102
Institutions	4241
Sources	1267
Authors	11,532
Conference papers	1893
Journal articles	877
Review papers	39
Total citations	6816
Average citations per document	2.43

3.2. Publication Maturity and Document Type Composition

Figure 2 shows the document type composition of the bibliometric corpus. Conference papers account for 67.4% of the records, while journal articles account for 31.2% and review papers account for only 1.39%. This distribution indicates that RAG remains a fast-moving technical field, dominated by rapid experimentation, system proposals, and application-driven implementations. The small proportion of review papers suggests that conceptual consolidation remains limited relative to the volume of technical production. This justifies the need for critical surveys that go beyond descriptive categorization and identify deeper assumptions, gaps, and research agendas.

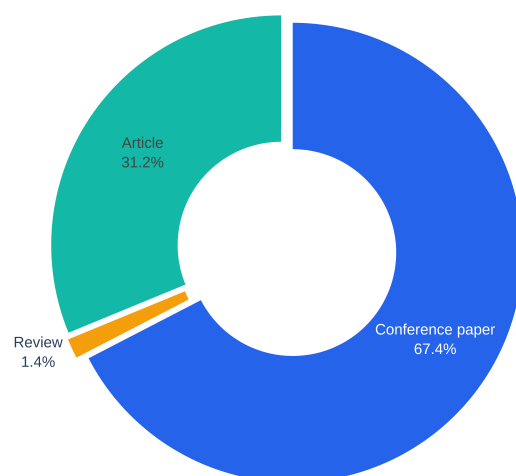


Figure 2. Document type composition of the RAG bibliometric corpus. The dominance of conference papers indicates a rapidly evolving technical field, while the small proportion of reviews suggests that conceptual consolidation remains limited.

3.3. Dominant Research Vocabulary

Figure 3 compares the most frequent author keywords and Keywords Plus. Author keywords are dominated by retrieval-augmented generation (1927), large language models (810), retrieval-augmented generation (RAG) (532), large language model (399), RAG (379), knowledge graph (150), artificial intelligence (133), natural language processing (133), and information retrieval (104). Keywords Plus similarly emphasizes retrieval-augmented generation (2439), language model (1,831), large language model (1567), information retrieval (1195), search engines (647), artificial intelligence (643), semantics (553), and question answering (482).

This distribution confirms that RAG research is organized around LLM grounding, retrieval mechanisms, semantic search, information retrieval, question answering, and knowledge graphs. However, terms related to curated thematic paths, canonical evidence units, thematic corpora, scriptures, legal taxonomies, educational taxonomies, or path-guided evidence pruning do not appear among the dominant keyword categories. This absence is important because many high-value domains already contain expert-authored thematic organization. The bibliometric vocabulary therefore supports the central motivation of this paper: current RAG research has not yet developed a consolidated retrieval framework for curated thematic corpora.

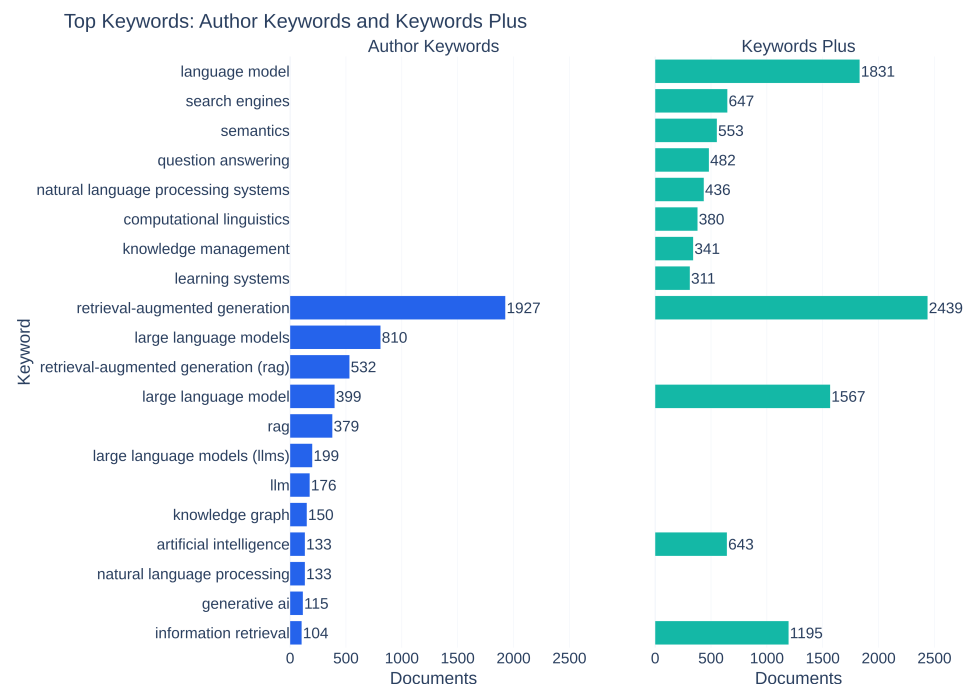


Figure 3. Top author keywords and Keywords Plus extracted from the Scopus RAG corpus. The dominant vocabulary centers on RAG, language models, information retrieval, search engines, semantics, question answering, artificial intelligence, and knowledge graphs but does not foreground curated thematic paths or canonical evidence units.

3.4. Thematic Evolution

Figure 4 shows thematic evolution over time. In 2023, knowledge graphs appear as an early theme. In 2024, retrieval-augmented generation, language model, large language model, and question answering become visible. In 2025, the field expands substantially, with retrieval-augmented generation, language model, large language model, information retrieval, search engines, artificial intelligence, semantics, natural language processing systems, computational linguistics, and knowledge management becoming central terms. The 2026 results should be interpreted cautiously because the dataset was exported in May 2026 and therefore represents only a partial publication year.

The thematic evolution indicates that graph-related retrieval has been present in the RAG literature, but the field has rapidly broadened toward general LLM-retrieval integration, information retrieval, and semantic search. It does not reveal a specialized cluster around curated thematic paths or canonical evidence pruning. This supports the argument that current graph and path-based RAG methods require further specialization for domains in which thematic organization already exists before indexing.



Figure 4. Theme evolution in the RAG bibliometric corpus from 2023 to 2026. The 2026 result should be interpreted as a partial-year signal because the export was collected in May 2026.

3.5. Geographical Distribution of Dominant Keywords

The country–keyword heatmap in Figure 5 indicates that China, the United States, and India are the leading contributors across the dominant RAG vocabulary. China has the highest keyword occurrences for retrieval-augmented generation, language model, large language model, and information retrieval. The United States and India also show strong activity across these terms. South Korea, Germany, and Japan contribute visible but smaller volumes.

The heatmap suggests that the core vocabulary of RAG research is globally shared, with major countries converging around LLMs, retrieval, information retrieval, search engines, and question answering. However, the same distribution does not show evidence of a specialized vocabulary around thematic path retrieval, curated evidence structures, or canonical-unit pruning. This strengthens the interpretation that ThemePath-RAG addresses a structural gap, not merely a gap in one national or regional research community.

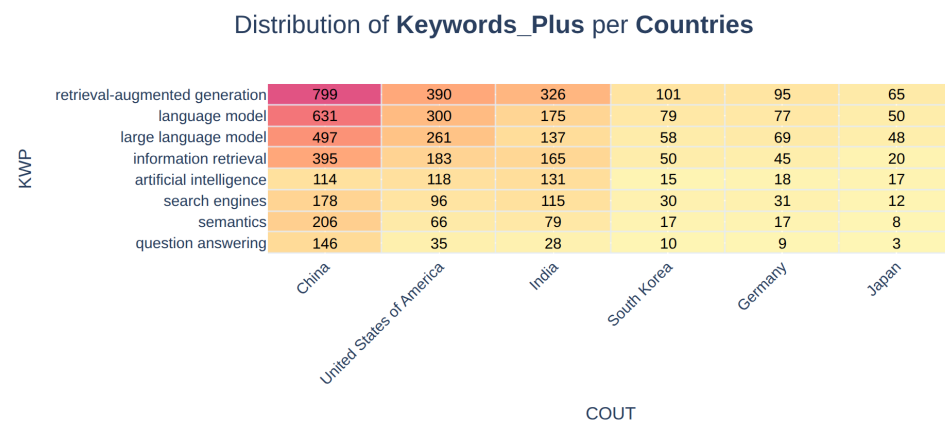


Figure 5. Relationship heatmap between leading countries and dominant Keywords Plus extracted from the Scopus RAG corpus. The same generic RAG vocabulary appears across major countries, suggesting global convergence around LLM-retrieval integration rather than curated thematic retrieval.

3.6. Bibliometric Gap Synthesis

Taken together, the bibliometric findings provide empirical support for the critical argument of this survey. First, the dominance of conference papers shows that RAG is still rapidly evolving, with many technical proposals but relatively few consolidating reviews. Second, keyword and theme evolution analyses show that the field is centered on LLMs, retrieval, information retrieval, question answering, semantics, and knowledge graphs. Third, the appearance of medical, legal, educational, compliance-oriented, and software-related terms indicates that RAG is increasingly applied to high-value knowledge domains. However, the bibliometric results do not reveal a mature cluster focused on curated thematic corpora, canonical evidence units, or thematic path-guided retrieval. This gap motivates ThemePath-RAG as a research agenda for exploiting expert-authored thematic paths and pruning evidence before grounded answer generation.

The bibliometric evidence listed in Table 2 shows that RAG research has expanded rapidly and is now distributed across general LLM-retrieval methods and multiple high-value domains. However, the dominant vocabulary and thematic clusters remain centered on generic retrieval, language models, information retrieval, semantics, question answering, knowledge graphs, and domain applications. This suggests that the field has not yet fully examined how RAG should operate when the retrieval corpus already contains expert-authored thematic paths. The next section therefore shifts from bibliometric mapping to a critical taxonomy of RAG methods based on their assumptions about knowledge representation. This is summarized in Figure 6.

Table 2. Bibliometric Evidence Supporting the ThemePath-RAG Research Gap.

Bibliometric Finding	Interpretation	Implication for ThemePath-RAG
Conference papers dominate the corpus (67.4%)	RAG remains a rapidly evolving technical field	Conceptual consolidation and critical surveys are still needed
Review papers account for only 1.39%	Survey literature is limited relative to technical production	A critical survey with a new research agenda is justified
Top keywords emphasize RAG, LLMs, information retrieval, search engines, semantics, and question answering	The field is organized around generic LLM-retrieval integration	Existing vocabulary does not foreground curated thematic paths
Knowledge graph appears among major keywords but with lower frequency than general RAG and LLM terms	Graph RAG is visible but not dominant	Graph-based retrieval needs more specialized analysis
Medical, legal, education, and compliance-oriented terms appear in the bibliometric landscape	RAG is increasingly applied to high-value structured domains	These domains may benefit from retrieval over curated thematic structures
Theme evolution shows expansion from knowledge graphs to broader RAG and LLM themes	The field has broadened quickly but remains generic	Pre-existing thematic organization remains under-theorized
No dominant cluster for curated thematic paths or canonical evidence units	Current RAG research has not consolidated around this problem	ThemePath-RAG addresses this missing retrieval setting

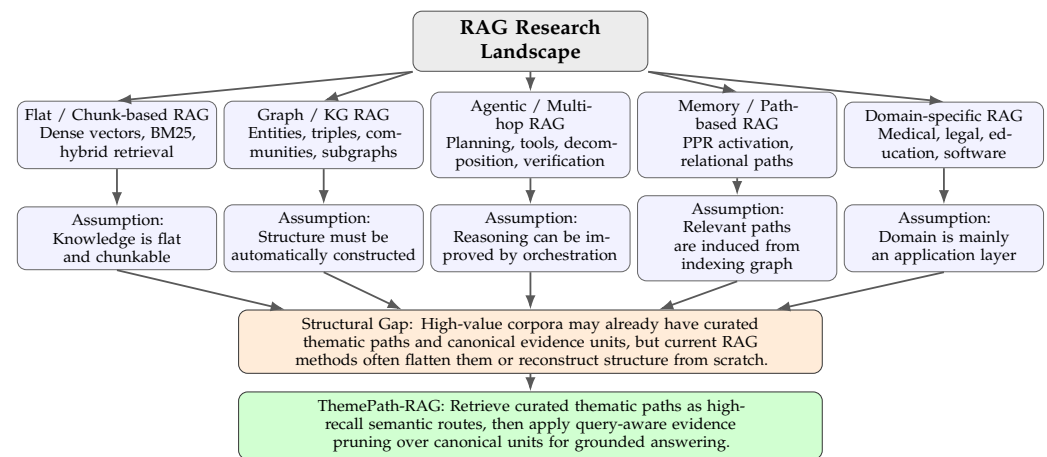


Figure 6. Critical taxonomy of RAG methods according to their assumptions about knowledge representation before retrieval. The upper category and assumption boxes are displayed in neutral light tones. The Structural Gap box is highlighted in orange to emphasize the identified limitation in existing RAG paradigms, whereas the ThemePath-RAG box is highlighted in green to indicate the proposed framework. ThemePath-RAG addresses the overlooked case in which the corpus already contains expert-authored thematic paths and canonical evidence units.

4. Critical Taxonomy of RAG Methods

4.1. Representative Scopus Literature Coverage

To avoid limiting the critical synthesis to a small number of well-known RAG papers, this survey additionally cites a representative set of Scopus-indexed RAG studies selected from the 2809 cleaned records. The selection was guided by title, abstract, author keyword, and Keywords Plus screening, with priority given to papers that represent distinct retrieval architectures, evaluation concerns, and application domains. The goal is not to claim that these are the only important papers in the corpus but to ensure that the proposed gap is grounded in a broad cross-section of the recent RAG literature. The selected references cover the following streams:

- The survey and review-oriented RAG studies are represented by [18–26].
- The flat, dense, lexical, and hybrid retrieval studies are represented by [27–38].
- The chunking, indexing, and structure-aware preprocessing studies are represented by [39–48].
- The graph-based, knowledge-graph, ontology, and hypergraph RAG studies are represented by [49–65].
- The multi-hop reasoning, evidence synthesis, and planning-oriented RAG studies are represented by [66–77].
- The agentic, multi-agent, tool-augmented, and workflow-oriented RAG studies are represented by [78–87].
- The evaluation, benchmark, hallucination, faithfulness, and reliability studies are represented by [88–101].
- The medical, biomedical, clinical, and health-oriented RAG studies are represented by [102–113].
- The legal, regulatory, compliance, tax, policy, and audit-oriented RAG studies are represented by [114–123].
- The education, learning, tutoring, and pedagogical RAG studies are represented by [124–133].
- The software engineering, code, repository, and requirements-oriented RAG studies are represented by [134–141].

- The multimodal, visual, image, video, and geospatial RAG studies are represented by [142–149].
- The security, privacy, cybersecurity, and IoT-oriented RAG studies are represented by [150–156].
- The religious, canonical, taxonomy, and faith-aligned RAG studies are represented by [157–161].

This study should also be distinguished from AI-augmented bibliometric systems that use agentic AI, full-paper retrieval, and retrieval-augmented synthesis to support dynamic scientometric analysis. For example, Bara and Oprea [162] propose an agentic framework that translates natural-language analytical requests into executable bibliometric workflows, retrieves and synthesizes full papers, and generates dynamic research reports. In contrast, the present study does not apply RAG to analyse the retrieved RAG literature corpus. Instead, Scopus metadata are used to provide bibliometric context for a critical survey, while ThemePath-RAG is proposed as a retrieval framework for external curated thematic corpora containing expert-authored paths and citeable canonical evidence units. Thus, the two studies operate at different levels: AI-augmented bibliometric frameworks support research-literature analysis, whereas ThemePath-RAG addresses evidence retrieval and pruning within structured domain corpora.

4.2. Flat Chunk-Based RAG

Flat chunk-based RAG is the most common paradigm. Documents are segmented into fixed or semantically informed chunks, embedded into vector space, and retrieved by dense similarity search. Sparse lexical methods such as BM25 may also be used, either independently or in hybrid form [163]. The central strength of this paradigm is simplicity. It can be applied to heterogeneous corpora without requiring domain-specific structure. Vector databases, approximate nearest neighbor search, and rerankers make this approach scalable and deployable.

However, flat chunk retrieval assumes that knowledge can be decomposed into independent retrieval units. This assumption is weak for corpora in which meaning depends on hierarchy, cross-reference, commentary, or thematic organization. A chunk may retrieve the correct passage but fail to expose its relationship to a broader theme. Conversely, a chunk may be semantically similar to the query but not authoritative for the answer. In canonical domains, such as legal codes or religious scripture, flat retrieval can also weaken citation precision because a retrieved chunk may mix primary text, commentary, and interpretation.

4.3. Sparse and Hybrid RAG

Sparse retrieval remains important in RAG because exact terms, identifiers, named entities, article numbers, verse numbers, medical codes, and legal references often carry high relevance. BM25 and related lexical methods are particularly useful in domains where exact matching is necessary. Hybrid retrieval combines sparse and dense retrieval to improve robustness against both lexical mismatch and semantic drift.

Despite these advantages, sparse and hybrid RAG usually operate over flat units. They can rank evidence more precisely than dense retrieval alone, but they do not inherently exploit curated thematic paths. In ThemePath-RAG, lexical matching is not treated as a complete retrieval solution. Instead, it becomes a query-aware evidence pruning mechanism after thematic path retrieval has narrowed the candidate space.

4.4. Graph-Based RAG

Graph-based RAG addresses the limitations of flat retrieval by representing entities, concepts, documents, or passages as nodes and their relationships as edges. GraphRAG,

LightRAG, and related approaches use graph structure to support local and global retrieval, summarization, and multi-hop reasoning [5,6]. This is a significant improvement over flat chunk retrieval because it makes relationships explicit.

However, graph-based RAG has two limitations. First, many methods automatically construct graphs from text. This is useful for open-domain corpora but may be redundant or noisy when expert-authored structures already exist. Second, graph retrieval often retrieves broad neighborhoods or communities. This can improve recall but may reduce evidence precision and increase prompt length. For high-stakes domains, broad graph context can make grounding less reliable because thematically related evidence may not be directly responsive to the user query.

4.5. *Agentic and Multi-Hop RAG*

Agentic RAG methods use planning, decomposition, tools, memory, verification, and iterative retrieval to answer complex questions [164,165]. These methods are especially useful when the query requires multiple retrieval steps or external tool use. Multi-hop RAG improves reasoning by retrieving evidence from multiple sources and connecting facts across documents.

Agentic RAG addresses reasoning orchestration, but it does not necessarily solve the representation problem. An agent can still retrieve redundant context if the underlying corpus is represented as flat chunks or broad graph neighborhoods. ThemePath-RAG is complementary to agentic RAG: an agent may use ThemePath-RAG as a retrieval tool when the domain contains curated thematic paths.

4.6. *Memory-Inspired and Path-Based RAG*

HippoRAG and HippoRAG 2 represent an important step toward associative retrieval. They use knowledge graphs and PPR to activate related nodes, aiming to mimic long-term memory and support multi-hop association [10,12]. PathRAG further argues that retrieving all graph neighbors or communities introduces redundancy, so it retrieves key relational paths instead [9].

These methods are directly relevant to ThemePath-RAG, but they differ in their assumptions. HippoRAG assumes that a graph must be built from passages using OpenIE and then activated by query-linked nodes. PathRAG assumes that relational paths should be extracted from an indexing graph. ThemePath-RAG instead begins from curated thematic paths that are already part of the domain knowledge organization. It adopts the anti-redundancy insight of PathRAG and the activation-oriented intuition of HippoRAG, but applies them to pre-structured thematic corpora.

4.7. *Domain-Specific RAG*

Domain-specific RAG has been applied to medicine, law, education, software engineering, finance, cybersecurity, policy, and many other areas. These works demonstrate the practical utility of RAG but often treat the domain as an application layer rather than a structural retrieval problem. For example, a legal RAG system may retrieve legal articles, but it may not explicitly exploit the hierarchical structure of books, chapters, sections, and clauses. A medical RAG system may retrieve guideline passages, but it may not explicitly use disease-treatment pathways as retrieval routes. Similarly, a Qur'anic QA system may retrieve verses, tafsir, and hadith, but it may not treat thematic paths as first-class retrieval objects.

This paper argues that a new subproblem is needed: RAG for curated thematic corpora. Such corpora require retrieval systems that preserve expert-authored structure while pruning evidence according to query-specific needs.

5. Structural Gap: Curated Thematic Corpora

5.1. Definition

We define a curated thematic corpus as a knowledge collection satisfying five conditions. First, it contains canonical evidence units, such as verses, articles, clauses, guidelines, learning outcomes, or policy items. Second, it is organized by human-authored themes, sub-themes, categories, or taxonomies. Third, it contains meaningful hierarchical or associative paths between broad concepts and specific evidence units. Fourth, evidence units may be connected to multiple themes. Fifth, answers require grounded citation to canonical units.

Formally, a curated thematic corpus can be represented as

$$C = (T, P, E, R, S), \quad (1)$$

where T is the set of thematic nodes, P is the set of curated thematic paths, E is the set of canonical evidence units, R is the set of relations between themes and evidence, and S is the set of source or citation metadata.

5.2. Examples

Examples of curated thematic corpora include religious scriptures, legal corpora, medical guidelines, educational curricula, library taxonomies, and policy documents. In religious scripture, themes may include stories, prophets, legal rulings, moral concepts, eschatology, and worship. In legal corpora, themes may map to books, chapters, sections, articles, and clauses. In medical guidelines, themes may map to diseases, diagnostic criteria, treatment pathways, contraindications, and evidence grades. In education, they may map to competencies, topics, learning outcomes, assessment items, and learning materials.

These corpora are not merely text collections. Rather, they are knowledge organizations. Flattening them into chunks can destroy useful structure, while reconstructing new graphs can ignore or distort expert curation.

5.3. Why Existing RAG Assumptions Are Insufficient

Existing RAG assumptions become insufficient in curated thematic corpora for three reasons. First, flat retrieval ignores domain-authored structure. Second, automatic graph construction may duplicate or corrupt existing expert organization. Third, graph-neighborhood retrieval may produce high recall but low precision by returning all evidence connected to a broad theme. The challenge is not simply to retrieve more context. The challenge is to exploit existing thematic structure and select only the evidence needed for a specific query.

The central gap can be stated as follows: existing RAG methods generally assume either flat text chunks or automatically constructed indexing graphs. However, many high-value corpora, including religious scriptures, legal codes, medical guidelines, and educational taxonomies, already possess curated thematic structures. Current graph-based RAG methods do not fully exploit such pre-existing thematic paths and often retrieve redundant node neighborhoods or overly broad communities.

6. Comparison with Existing Methods

Table 3 shows that ThemePath-RAG differs from existing RAG paradigms primarily in its treatment of structure source and retrieval unit. The framework is not defined by a new embedding model, a new lexical ranking function, or a replacement for graph retrieval. Instead, it is designed for a specific corpus condition: the domain already provides expert-authored thematic paths that connect broader concepts to citeable canonical evidence units. Under this condition, the central retrieval problem is not only finding relevant text, but

preserving the existing thematic organization while selecting the small subset of evidence that directly addresses a user query.

Table 3. Comparison Between ThemePath-RAG and Existing RAG Paradigms.

Method	Retrieval Unit	Structure Source	Main Limitation	ThemePath-RAG Difference
Vector RAG	Text chunks	Chunking and embedding	Ignores thematic hierarchy	Retrieves curated thematic paths first
BM25 RAG	Terms and documents	Lexical index	Weak semantic structure	Uses lexical scoring for evidence pruning
Hybrid RAG	Chunks	Dense and sparse indexes	Still mostly flat	Adds path-level retrieval and pruning
GraphRAG	Communities or sub-graphs	Auto-generated graph	Broad context and redundancy	Uses curated paths and pruned canonical evidence
LightRAG	Local/global graph retrieval	Auto-generated graph	May retrieve noisy neighbors	Avoids all-neighbor retrieval
HippoRAG	KG nodes and passages	OpenIE graph	Built for constructed KGs	Uses curated thematic paths instead of rebuilding graph from scratch
PathRAG	Relational paths	Indexing graph	General graph paths, not domain-authored thematic paths	Treats expert-authored thematic paths as retrieval routes
ThemePath-RAG	Thematic paths and evidence units	Curated thematic corpus	Requires curated structure	Designed for pre-structured canonical corpora

Vector RAG, BM25 RAG, and Hybrid RAG generally retrieve from flat evidence units such as chunks, passages, documents, or individual records. These methods remain important components of ThemePath-RAG rather than competing alternatives that must be discarded. Dense retrieval can be used to identify candidate thematic paths, BM25 can support exact lexical matching among candidate evidence units, and hybrid scoring can combine complementary retrieval signals. However, when these methods operate only over a flat index, they do not explicitly use the thematic route through which a canonical evidence unit is connected to a broader domain concept. ThemePath-RAG introduces this path-level stage before final evidence ranking: it first retrieves relevant curated paths, then applies lexical and semantic scoring only after the candidate space has been constrained by those paths.

GraphRAG and LightRAG address limitations of flat retrieval by representing entities, passages, concepts, and relationships as graph structures. Their graph-based retrieval mechanisms can support multi-hop reasoning and make relationships between information units more explicit. Nevertheless, many graph-based RAG approaches construct indexing graphs automatically from unstructured documents through entity extraction, relation extraction, summarization, or community detection. ThemePath-RAG addresses a different setting in which the relevant structure already exists before indexing. It does not attempt to replace automatically constructed graphs in open-domain corpora; rather, it uses pre-existing thematic paths as retrieval routes in corpora where those paths have been authored, maintained, or validated by domain experts.

The comparison with GraphRAG and LightRAG should therefore not be interpreted as a claim that graph retrieval necessarily produces redundant context in every application. Broad graph neighborhoods, local neighbors, or community-level retrieval may be beneficial when the query requires high recall or multi-hop synthesis. However, in

citation-sensitive corpora, a retrieved neighborhood or endpoint theme can contain more evidence than is needed to answer a specific question. ThemePath-RAG treats the retrieved thematic path as a high-recall candidate space rather than final context. The subsequent evidence-scoring and global-pruning stages are intended to select the most query-responsive canonical units from that space.

HippoRAG and PathRAG are especially relevant because both recognize that retrieval should move beyond isolated flat passages. HippoRAG uses constructed knowledge graphs and Personalized PageRank-based activation to retrieve associated passages, whereas PathRAG retrieves relational paths from an indexing graph to reduce redundancy in graph-based retrieval. ThemePath-RAG shares their broader intuition that structural retrieval can improve evidence selection. Its distinction lies in the origin and meaning of the paths. HippoRAG begins from graph structures induced from textual knowledge, and PathRAG operates over relational paths extracted from an indexing graph. ThemePath-RAG begins from thematic paths that are already part of the domain's knowledge organization and uses those paths to guide candidate evidence expansion and canonical-unit pruning.

Ontology-based and hierarchical retrieval methods are also related because they can preserve taxonomies, concept hierarchies, controlled vocabularies, or semantic relationships. The difference is not that such methods lack structure but that ThemePath-RAG explicitly separates two decisions: selecting a meaningful expert-authored route and selecting the final evidence units beneath or around that route. In this formulation, a hierarchy or ontology is not treated only as metadata for semantic matching or navigation. It becomes an interpretable retrieval object that constrains a subsequent query-aware evidence-pruning process. This distinction is particularly relevant when a domain requires direct citation to verses, legal clauses, clinical recommendations, learning outcomes, or policy provisions.

ThemePath-RAG also has a clear boundary condition. It is not intended as a universal replacement for Vector RAG, GraphRAG, or ontology-based RAG. The framework is appropriate only when a corpus contains meaningful curated paths, when those paths connect to citeable evidence units, and when broad thematic retrieval must be refined into a compact evidence set. In corpora without reliable curated structure, automatic graph construction, flat vector retrieval, lexical retrieval, or hybrid retrieval may be more appropriate. The Qur'anic QA proof of concept in this study evaluates this architecture in one curated thematic corpus; direct empirical comparison with GraphRAG, HippoRAG, PathRAG, ontology-based retrieval, and additional baselines remains necessary in future language-matched benchmarks.

7. ThemePath-RAG Framework

7.1. Conceptual Overview

ThemePath-RAG is designed for corpora that already contain curated thematic paths. It does not replace vector retrieval, BM25, or graph retrieval. Instead, it organizes them into a staged architecture. The core pipeline (see Figure 7) is query rewriting, thematic path retrieval, candidate evidence expansion, query-aware evidence scoring, evidence pruning, and citation-grounded answer generation.

The framework can be summarized as

$$q \rightarrow P_q \rightarrow E(P_q) \rightarrow E_q^* \rightarrow A(q, E_q^*), \quad (2)$$

where q is the user query, P_q is the retrieved thematic path set, $E(P_q)$ is the candidate evidence space expanded from the paths, E_q^* is the pruned evidence subset, and A is the generated answer.

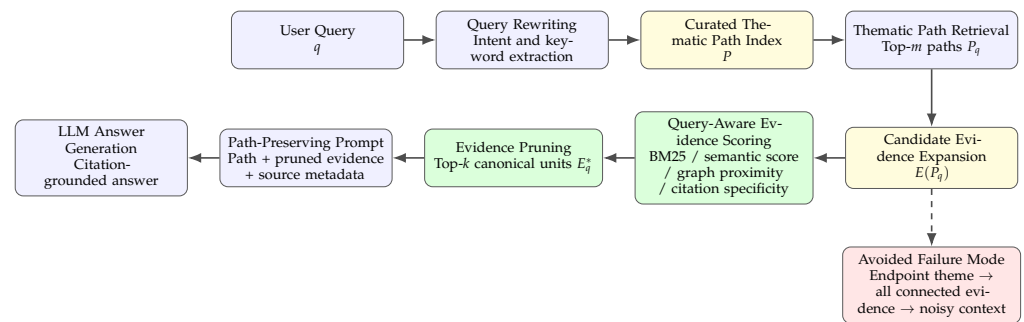


Figure 7. ThemePath-RAG pipeline. Light purple boxes indicate the query, prompt, and answer-generation stages; light yellow boxes indicate thematic-index retrieval and candidate-evidence expansion; green boxes indicate query-aware scoring and evidence pruning; and the light red box highlights the avoided failure mode. Solid arrows show the ordered retrieval-and-generation workflow, whereas the dashed arrow indicates that endpoint-only expansion without pruning may produce noisy context. Curated thematic paths are used as high-recall semantic routes, but endpoint themes are treated only as candidate spaces. Query-aware evidence pruning selects a compact set of canonical evidence units before grounded answer generation.

7.2. Stage 1: Thematic Path Retrieval

The first stage retrieves one or more thematic paths relevant to the query. Each thematic path is represented as a textual sequence, for example, Story → Prophets → Miracle of Prophets → Prophet Muhammad → Isra’ Mi’raj. The path can be indexed using dense embeddings, BM25, or hybrid retrieval. The goal is high recall. At this stage, it is acceptable to retrieve multiple candidate paths because evidence pruning will refine the context later.

7.3. Stage 2: Candidate Evidence Expansion

For each retrieved path, the system expands candidate evidence from the endpoint theme, intermediate themes, cross-linked subthemes, canonical units connected to the themes, and commentary or explanatory nodes associated with the units. This differs from endpoint-only retrieval. If only the endpoint is used, the system may miss evidence from intermediate or related thematic nodes. Conversely, if all connected evidence is sent to the LLM, the system may retrieve too much. Candidate expansion is therefore followed by pruning.

7.4. Stage 3: Query-Aware Evidence Scoring

Each candidate evidence unit is scored according to query relevance and path context. We propose the following scoring function:

$$\begin{aligned} \text{Score}(e, q, p) = & \alpha \text{LexSim}(q, e) + \beta \text{SemSim}(q, e) \\ & + \gamma \text{PathSim}(q, p) + \delta \text{GraphProx}(e, p) \\ & + \eta \text{SourcePriority}(e) + \lambda \text{CitationSpecificity}(e), \end{aligned} \quad (3)$$

where e is a candidate evidence unit, q is the user query, p is a retrieved thematic path, $\text{LexSim}(q, e)$ is lexical similarity such as BM25, $\text{SemSim}(q, e)$ is optional semantic similarity, $\text{PathSim}(q, p)$ is similarity between the query and thematic path, $\text{GraphProx}(e, p)$ measures graph proximity between evidence and path nodes, $\text{SourcePriority}(e)$ prioritizes primary sources, and $\text{CitationSpecificity}(e)$ rewards evidence with precise citation metadata.

The scoring function is intentionally modular. In a lightweight implementation, LexSim , SemSim , and PathSim may be sufficient. In a more advanced implementation, graph proximity, source reliability, and citation specificity can be added.

7.5. Stage 4: Evidence Pruning

Evidence pruning selects the top- k evidence units after scoring as follows:

$$E_q^* = \text{TopK}(E(P_q), \text{Score}(e, q, p)). \quad (4)$$

More advanced pruning can enforce diversity constraints, source constraints, or path coverage constraints. For example, a religious QA system may select top Qur'anic verses first, then tafsir passages, then hadith if relevant. A legal QA system may select statutory articles before commentary. A medical QA system may prioritize guideline statements with evidence grades.

7.6. Stage 5: Citation-Grounded Answer Generation

The final LLM prompt should preserve path structure. Instead of concatenating evidence as a flat list, ThemePath-RAG organizes evidence as the user query, retrieved thematic path, selected evidence units, source metadata, and answer instructions. This prompt structure preserves both thematic context and evidence precision. It also supports interpretability because the answer can be traced from query to thematic path to pruned evidence.

7.7. Proof-of-Concept Instantiation

The Qur'an is a canonical corpus with multiple layers of organization, including surah and ayah structure; themes and subthemes; stories of prophets and communities; legal, ethical, theological, and eschatological concepts; tafsir and interpretive traditions; hadith links; and historical context. A Qur'anic RAG system must not merely generate a plausible answer. It must ground claims in precise evidence, identify relevant ayah, distinguish translation from tafsir, and avoid unsupported religious claims. Recent RAG studies on Qur'anic and religious texts show growing interest in retrieval-grounded sacred-text dialogue and low-resource religious QA, including Qur'anic QA, Seerah QA, Bhagavad Gita QA, and religious Punjabi texts [157–160,166].

A recent benchmark of RAG for Qur'anic studies evaluated 13 open-source LLMs across human-assessed context relevance, answer faithfulness, and answer relevance using a descriptive dataset of Qur'anic surahs [13]. The study showed that retrieval grounding is important for sensitive religious QA and that model scale alone does not fully determine faithfulness. However, it also motivates a retrieval-level question that remains unresolved: how should the system select the most query-relevant ayah, tafsir, or hadith evidence when the broader thematic category is already known?

The limitation of endpoint-only thematic retrieval is clear in this setting. A thematic retriever may identify the correct endpoint node, for example, Isra' Mi'raj. However, if all verses connected to this node are retrieved, the LLM may receive too much evidence. Some verses may be broadly related to the Prophet Muhammad, revelation, prayer, miracles, or night journey but not directly answer the user query. This creates context noise, citation dilution, and token inefficiency. ThemePath-RAG addresses this by separating theme discovery from evidence selection. It retrieves a thematic path, expands candidate ayah, tafsir, and hadith from path nodes, scores evidence using query-aware criteria, prunes evidence to a compact set, and generates an answer with citation metadata. For example, the path Story → Prophets → Miracle of Prophets → Prophet Muhammad → Isra' Mi'raj provides a semantic route, but the final answer may require different evidence depending on whether the query asks about Qur'anic evidence, prayer obligation, physical or spiritual interpretation, or historical context.

The full ThemePath-RAG scoring function in Equation (3) is intended as a modular framework for curated thematic corpora. To operationalize the framework in an

initial Qur'anic question-answering proof of concept, we implemented a simplified evidence-scoring function based on lexical similarity, semantic similarity, and thematic-path similarity:

$$Score_{PoC}(e, q, p) = \alpha \widetilde{LexSim}(q, e) + \beta \widetilde{SemSim}(q, e) + \gamma \widetilde{PathSim}(q, p), \quad (5)$$

where e is a candidate ayat, q is the user query, and p is a retrieved thematic path. The terms $\widetilde{LexSim}(q, e)$, $\widetilde{SemSim}(q, e)$, and $\widetilde{PathSim}(q, p)$ denote normalized lexical, semantic, and path-similarity scores, respectively. The non-negative weights α , β , and γ satisfy

$$\alpha + \beta + \gamma = 1. \quad (6)$$

This formulation allows the contribution of lexical relevance, semantic correspondence, and thematic-path alignment to be varied without changing the retrieval pipeline.

For each fixed English query q , an English thematic-path vector-search stage first produced the top $m = 3$ candidate thematic paths. A Cypher-based graph retrieval procedure then expanded the ayat associated with these candidate paths. The explicit path-similarity component was computed as cosine similarity between the OpenAI embedding of the English query and the embedding of each candidate path's textual representation.

Lexical similarity was calculated using BM25 over the unique English ayat texts in the candidate set for each query. Constructing the BM25 corpus from unique ayat texts prevented repeated occurrences of the same ayat under multiple thematic paths from altering the lexical scoring statistics. The resulting BM25 score was then assigned to every corresponding (e, p) candidate pair.

Semantic similarity was computed using OpenAI text-embedding-3-small embeddings for the English query and candidate ayat texts. Candidate ayat embeddings were temporarily indexed in an in-memory Chroma collection configured with cosine distance. Semantic similarity was calculated as $1 - \text{distance}$ and bounded to the interval $[0, 1]$. Path similarity was computed directly using cosine similarity between query and path embeddings.

For each query, raw lexical, semantic, and path-similarity values were independently min-max normalized across all expanded (e, p) candidate pairs. The normalized component values were then combined using Equation (5). Duplicate ayat occurring through multiple thematic paths were collapsed by canonical ayat reference, retaining the occurrence with the highest combined score. The resulting unique ayat were ranked globally, and the top $k = 3$ were retained as the final retrieval contexts.

For the proof-of-concept implementation, the similarity weights were specified heuristically as follows:

$$(\alpha, \beta, \gamma) = (0.30, 0.50, 0.20), \quad (7)$$

This configuration assigns the greatest weight to semantic similarity, while lexical similarity and thematic-path similarity provide complementary retrieval signals. The weights were selected to emphasize semantic correspondence between the user query and candidate ayat; they are not claimed to represent an optimized weighting scheme. Systematic tuning of alternative weight combinations using a separate validation set is left for future work.

Algorithm 1 presents the general retrieval and evidence-pruning procedure used in the Qur'anic QA prototype. The algorithm evaluates the retrieval component of ThemePath-RAG only. It does not by itself evaluate answer-generation quality, answer faithfulness, citation correctness, or theological interpretation.

Algorithm 1 ThemePath-RAG Retrieval and Context Relevance Evaluation

Require: User query q ; thematic path index $\mathcal{P}$; thematic-path-to-ayah mapping $\mathcal{E}(\cdot)$; lexical similarity function $LexSim$; semantic similarity function $SemSim$; path similarity function $PathSim$; path budget m ; evidence budget k ; weights α , β , and γ , where $\alpha + \beta + \gamma = 1$

Ensure: Retrieved contexts $\mathcal{C}_q$; context relevance score $CR(q, \mathcal{C}_q)$

```

1:  $\mathcal{T} \leftarrow \emptyset$ 
2:  $\mathcal{S} \leftarrow \emptyset$ 
3: for all thematic paths  $p \in \mathcal{P}$  do
4:    $s_p^{raw} \leftarrow PathSim(q, p)$  ▷ Compute query-to-path similarity
5:    $\mathcal{T} \leftarrow \mathcal{T} \cup \{(p, s_p^{raw})\}$ 
6: end for
7:  $\mathcal{P}_q \leftarrow TopM(\mathcal{T}, m)$  ▷ Select the top  $m$  thematic paths
8: for all  $(p, s_p^{raw}) \in \mathcal{P}_q$  do
9:    $\mathcal{E}_p \leftarrow \mathcal{E}(p)$  ▷ Expand candidate ayat associated with path  $p$ 
10:  for all candidate ayat  $e \in \mathcal{E}_p$  do
11:     $s_{lex}^{raw} \leftarrow LexSim(q, e)$ 
12:     $s_{sem}^{raw} \leftarrow SemSim(q, e)$ 
13:     $\mathcal{S} \leftarrow \mathcal{S} \cup \{(e, p, s_{lex}^{raw}, s_{sem}^{raw}, s_p^{raw})\}$ 
14:  end for
15: end for
16: Normalize lexical, semantic, and path similarity scores in  $\mathcal{S}$  to obtain  $\tilde{s}_{lex}$ ,  $\tilde{s}_{sem}$ , and  $\tilde{s}_{path}$ 
17: for all  $(e, p, \tilde{s}_{lex}, \tilde{s}_{sem}, \tilde{s}_{path}) \in \mathcal{S}$  do
18:    $Score(e, q, p) \leftarrow \alpha \tilde{s}_{lex} + \beta \tilde{s}_{sem} + \gamma \tilde{s}_{path}$ 
19: end for
20:  $\mathcal{S}^* \leftarrow DeduplicateByAyahID(\mathcal{S})$  ▷ Retain the highest-scoring occurrence of each ayat
21:  $\mathcal{S}^\dagger \leftarrow SortDescending(\mathcal{S}^*, Score)$ 
22:  $\mathcal{R}_q \leftarrow TopK(\mathcal{S}^\dagger, k)$  ▷ Select the top  $k$  unique ayat globally
23:  $\mathcal{C}_q \leftarrow ExtractAyahTexts(\mathcal{R}_q)$ 
24:  $CR(q, \mathcal{C}_q) \leftarrow EvaluateContextRelevance(q, \mathcal{C}_q)$  ▷ RAGAS-based context relevance evaluation
25: return  $\mathcal{C}_q, CR(q, \mathcal{C}_q)$ 

```

The experiment used a fixed thematic-path retrieval budget of $m = 3$ and a final evidence-retrieval budget of $k = 3$. The same final top- k target was used for both the Vector RAG baseline and ThemePath-RAG to reduce differences attributable to context quantity.

For each selected ayat, the retrieval procedure retained its canonical identifier, associated thematic path, component similarity values, aggregated evidence score, and final ranking position. These retrieval attributes enable transparent inspection of thematic-path selection and evidence-pruning behaviour.

8. Proof-of-Concept Evaluation in Qur'anic Question Answering

8.1. Evaluation Objective

This proof-of-concept experiment was conducted to move ThemePath-RAG beyond a purely conceptual framework and to examine whether its retrieval stages can be operationalized over a curated Qur'anic thematic corpus. The experiment focused on retrieved-context relevance rather than answer generation, theological interpretation, or cross-domain generalization.

The primary research question was as follows:

How does the ThemePath-RAG prototype compare with a Vector RAG baseline in retrieving contexts relevant to Qur'anic question-answering queries?

The experiment should be interpreted as a retrieval-stage evaluation. It does not establish that ThemePath-RAG is superior to conventional Vector RAG, nor does it evaluate answer faithfulness, citation accuracy, doctrinal correctness, answer completeness, or expert acceptability. Its purpose is to examine the practical behaviour of thematic-path retrieval, candidate evidence expansion, weighted evidence scoring, and global evidence pruning.

8.2. Evaluation Protocol

The evaluation set comprised $N = 150$ human-created Indonesian Qur'anic question-answering instances. To construct the dataset, thematic paths and their associated ayat were randomly sampled from the curated thematic index by Indonesian Islamic scholars, (available at <https://alquranalhadi.com> (accessed on 21 June 2026)). Human annotators then formulated questions based on the sampled thematic paths and associated ayat, so that each question reflected a specific thematic context and canonical Qur'anic evidence. Exact duplicate questions and semantically near-duplicate questions were removed during dataset preparation.

During dataset preparation, each Indonesian question was translated once into English using gpt-4o-mini. The resulting fixed English translations were stored before retrieval and were passed unchanged to the ThemePath-RAG pipeline. Thus, translation was not repeated during the retrieval or evaluation loop. The same English query representation was used for thematic-path retrieval, candidate-ayah scoring, and RAGAS context-relevance evaluation of the retrieved English ayat contexts.

Vector RAG used the original Indonesian question and Indonesian ayat representations. Therefore, the two systems were paired by underlying question identity but were not fully language-controlled. This representation difference may affect both retrieval behaviour and LLM-based relevance assessment.

Context relevance was calculated using ContextRelevance from the RAGAS evaluation framework with gpt-4o-mini as the LLM evaluator. For each instance, the evaluator received only the query and the final retrieved contexts; no answer-generation model was invoked in this experiment. The evaluation therefore measures retrieval-stage context relevance rather than answer quality, faithfulness, citation correctness, or theological validity.

The paired design was selected to control for variation in question difficulty. For each question instance q_i , the final retrieved contexts from Vector RAG and ThemePath-RAG were evaluated using the same RAGAS-based context-relevance procedure. This enabled direct query-level comparison between the two retrieval settings.

The Vector RAG baseline treated individual Qur'anic ayat as independent retrieval units and returned the top $k = 3$ ayat for each question. ThemePath-RAG first ranked all curated thematic paths and retained the top $m = 3$ paths for candidate evidence expansion. Candidate ayat associated with the selected paths were scored using the weighted combination of lexical similarity, semantic similarity, and thematic-path similarity defined in Equation (5) with the weight defined in Equation (7). Candidate ayat were then deduplicated by ayat identifier, ranked globally, and pruned to the top $k = 3$ unique ayat.

Context relevance was evaluated using the RAGAS framework with gpt-4o-mini as the LLM evaluator [167]. Scores ranged from 0 to 1, where higher values indicate greater relevance between the user question and the retrieved contexts.

Table 4 summarizes the experimental configuration.

The final retrieval budget was closely matched between the two approaches. Vector RAG returned exactly three ayat for every question, while ThemePath-RAG returned three ayat for 149 questions and two ayat for one question. This design reduces the influence of context quantity and allows the comparison to focus on the relevance of the final evidence set.

However, the comparison is not a fully language-controlled benchmark. Vector RAG used Indonesian question and ayat representations, whereas ThemePath-RAG used English translations of the same questions and English ayat representations. This difference may affect both retrieval behaviour and LLM-based relevance assessment.

Table 4. Configuration of the Qur’anic QA proof-of-concept experiment.

Parameter	Vector RAG	ThemePath-RAG
Evaluation instances	150	150
Canonical evidence unit	Ayat	Ayat
Query representation	Indonesian	English translation
Ayat representation	Indonesian translation	English translation
Thematic-path representation	Not applicable	English thematic paths
Path-selection rule	Not applicable	Top- $m = 3$ paths
Lexical similarity weight	Not applicable	0.30
Semantic similarity weight	Not applicable	0.50
Path-similarity weight	Not applicable	0.20
Initial path-candidate generation	Not applicable	English thematic-path vector search
Candidate evidence expansion	Verse-level retrieval	Cypher-based retrieval from selected thematic paths
Lexical scoring corpus	Not applicable	Unique English candidate ayat texts per query
Semantic and path embedding model	text-embedding-3-small	text-embedding-3-small
Similarity function	Cosine similarity	Cosine similarity
Component-score normalization	Not applicable	Per-query min-max normalization across expanded (e, p) pairs
Duplicate handling	Not applicable	Maximum combined score retained per ayat reference
Final evidence selection	Top- $k = 3$ ayat	Global top- $k = 3$ unique ayat
Actual retrieved contexts	3 for all questions	3 for 149 questions; 2 for 1 question
Mean retrieved contexts	3.00	2.99
Mean context words	77.96	103.13
Evaluation framework	RAGAS	RAGAS
LLM evaluator	gpt-4o-mini	gpt-4o-mini
Primary metric	Context relevance	Context relevance

8.3. Context Relevance Results

Table 5 presents the descriptive results of the paired evaluation.

Table 5. Context relevance results for Vector RAG and ThemePath-RAG.

Measure	Vector RAG	ThemePath-RAG
Number of paired questions	150	150
Mean context relevance score	0.920	0.798
Standard deviation	0.201	0.285
Median context relevance score	1.000	1.000
First quartile	1.000	0.500
Third quartile	1.000	1.000
Minimum score	0.000	0.000
Maximum score	1.000	1.000
Perfect relevance score, 1.00	128 (85.3%)	94 (62.7%)
Zero relevance score, 0.00	1 (0.7%)	3 (2.0%)
Mean retrieved contexts	3.00	2.99
Mean context words	77.96	103.13

Vector RAG achieved a mean context relevance score of 0.920, whereas ThemePath-RAG achieved a mean score of 0.798. Vector RAG obtained a perfect relevance score on 128 of 150 questions (85.3%), compared with 94 questions (62.7%) for ThemePath-RAG. Conversely, ThemePath-RAG produced a zero relevance score on three questions (2.0%), whereas Vector RAG produced a zero score on one question (0.7%).

For each question q_i , the paired score difference was calculated as

$$\Delta_i = CR_{\text{ThemePath}}(q_i) - CR_{\text{Vector}}(q_i), \quad (8)$$

where $CR_{\text{ThemePath}}(q_i)$ and $CR_{\text{Vector}}(q_i)$ denote the context relevance scores of ThemePath-RAG and Vector RAG, respectively.

ThemePath-RAG obtained a higher score than Vector RAG on 11 questions (7.3%), produced the same score on 88 questions (58.7%), and obtained a lower score on 51 questions (34.0%). The mean paired difference was -0.122 , with a bootstrap 95% confidence interval of $[-0.173, -0.070]$. A paired Wilcoxon signed-rank test indicated that Vector RAG remained significantly stronger under the current configuration ($W = 429.5$, $p < 0.001$).

Therefore, the ThemePath-RAG prototype did not surpass the Vector RAG baseline in RAGAS context relevance. Nevertheless, the experiment demonstrates that ThemePath-RAG can be instantiated as a complete retrieval-and-pruning pipeline with a fixed final evidence budget.

8.4. Qualitative Retrieval Examples

Table 6 summarizes three illustrative retrieval cases. The examples were selected to show direct retrieval success, a verse-level retrieval failure, and a case in which thematic-path guidance produced a more focused evidence set than Vector RAG.

Table 6. Illustrative qualitative comparison of Vector RAG and ThemePath-RAG retrieval outcomes.

Question	Vector RAG	ThemePath-RAG	Interpretation
Bagaimana Musa menjelaskan siapa Tuhan itu kepada Fir'aun? (How did Moses explain who the Lord is to Pharaoh?)	1.00	1.00	Both methods retrieved evidence directly related to Fir'aun's question and Musa's explanation of the Lord. This case shows that thematic-path retrieval can preserve direct verse-level relevance when the selected paths align closely with the question.
Bagaimana jawaban Hud kepada kaumnya ketika ia dituduh kurang waras? (How did Hud respond to his people when they accused him of lacking reason?)	1.00	0.50	Vector RAG retrieved the verse in which Hud explicitly rejects the accusation. ThemePath-RAG retrieved related evidence concerning Hud's call and his people's rejection but did not retrieve the most direct verse containing his response.
Apa akibat yang disebutkan bagi orang-orang yang zalim? (What consequence is mentioned for the wrongdoers?)	0.50	1.00	ThemePath-RAG retrieved a more consistently consequence-oriented evidence set. The selected thematic route and final evidence ranking are detailed in Table 7.

The first two cases illustrate that ThemePath-RAG can retrieve direct canonical evidence when thematic-path selection is well aligned with the query, but it may still miss the most specific ayat for verse-oriented questions. The third case demonstrates how a curated thematic route can focus the candidate evidence space around the intended semantic interpretation.

Table 7 presents the complete retrieval trace for the third case. Rather than treating the thematic endpoint as final evidence, ThemePath-RAG uses the selected path to define a focused candidate space and then ranks canonical ayat globally using lexical, semantic, and thematic-path similarity. ThemePath-RAG first ranked the curated thematic paths and retained the top $m = 3$ paths. Candidate ayat associated with these paths were then expanded, scored using Equation (5), deduplicated by ayat identifier, and globally pruned to the top $k = 3$ ayat.

Table 7. Worked ThemePath-RAG retrieval trace for the question on consequences for wrongdoers. Path names are abbreviated for readability. Raw path similarity is used for path ranking; component values in Panel B are normalized before weighted aggregation.

Panel A. Top- $m = 3$ thematic-path retrieval						
Rank	Curated Thematic Path	Raw Path Similarity				
P1	Story $\rightarrow$ People $\rightarrow$ Ashhab al-Ukhdud $\rightarrow$ The Bad Consequences of Their Crime	0.539828				
P2	Morals $\rightarrow$ Destructive Morals $\rightarrow$ Tyranny $\rightarrow$ On the Day of Judgment, the Wrongdoers Will Be in Fear	0.444480				
P3	Sharia $\rightarrow$ Fiqh Mu'amalah $\rightarrow$ Condemnation for Those Who Cheat	0.414062				
Panel B. Global top- $k = 3$ evidence ranking after candidate expansion and deduplication						
Rank	Selected ayat and abbreviated evidence	Path	$\widetilde{LexSim}$	$\widetilde{SemSim}$	$\widetilde{PathSim}$	Score
1	Qur'an 42:22: wrongdoers are fearful of what they have earned, and the consequence will inevitably befall them	P2	1.000000	1.000000	0.241862	0.848372
2	Qur'an 85:10: those who commit crimes against believing men and women and do not repent face the punishment of Hell and the Burning Fire	P1	0.000000	0.933415	1.000000	0.666707
3	Qur'an 83:1: warning of woe for those who give less than due	P3	0.000000	0.902207	0.000000	0.451103
Panel C. Vector RAG top- $k = 3$ retrieval						
Rank	Retrieved ayat and abbreviated evidence	Vector score				
1	Qur'an 6:21: wrongdoers who fabricate falsehood or deny Allah's signs do not prosper	0.684167				
2	Qur'an 24:50: a statement identifying a group as wrongdoers, without directly describing their consequence	0.682845				
3	Qur'an 10:54: wrongdoers witness punishment and conceal their regret	0.682791				

Although P1 obtained the highest raw query-to-path similarity, the direct thematic route for the query was P2:

Morals → Destructive Morals → Tyranny → On the Day of Judgment, the Wrongdoers Will Be in Fear.

The candidate ayat from P2, Qur'an 42:22, received the highest final score because it achieved the maximum normalized lexical and semantic similarity values among the expanded candidates. This illustrates the intended role of thematic paths: they define a focused semantic candidate space, while final evidence selection remains dependent on query-aware scoring at the canonical ayat level.

Vector RAG retrieved one relevant consequence-oriented verse, Qur'an 10:54, but its higher-ranked evidence included a general statement that wrongdoers do not prosper (Qur'an 6:21) and a verse that identifies a group as wrongdoers without directly explaining their consequence (Qur'an 24:50). The resulting RAGAS context relevance score was 0.50. In contrast, ThemePath-RAG retrieved a more consistently consequence-oriented evidence set and obtained a context relevance score of 1.00.

For illustration, the selected ThemePath-RAG evidence supports the following citation-grounded response:

The Qur'an states that wrongdoers will be fearful of what they have earned and that the consequence of their deeds will inevitably befall them (Qur'an 42:22). It also warns that those who commit crimes against believing men and women and do not repent face the punishment of Hell and the Burning Fire (Qur'an 85:10). A further warning of woe is given to those who cheat in measure (Qur'an 83:1).

This illustrative response is included to demonstrate the complete retrieval-to-evidence pathway. Answer-generation quality, faithfulness, and citation correctness were not quantitatively evaluated in the present proof of concept.

8.5. Interpretation of the Proof-of-Concept Results

The proof-of-concept experiment demonstrates that ThemePath-RAG can be implemented over a curated Qur'anic thematic corpus and assessed using a retrieval-stage metric. This directly addresses the limitation of presenting ThemePath-RAG only as a conceptual framework.

However, the results do not support a claim that ThemePath-RAG currently outperforms Vector RAG in context relevance. Vector RAG achieved a higher mean relevance score, a greater proportion of perfect scores, and significantly stronger paired performance under the present evaluation configuration.

The result suggests that curated thematic paths can provide useful retrieval signals, but their effectiveness depends on accurate thematic-path ranking and sufficiently selective evidence pruning. ThemePath-RAG should therefore be interpreted as a feasible retrieval framework for curated thematic corpora rather than as a method whose superiority follows automatically from the use of thematic paths.

Several limitations should be considered. First, the Vector RAG and ThemePath-RAG settings used different language representations. Second, the weighting scheme was specified heuristically and was not optimized on an independent validation set. Third, the experiment did not use manually annotated gold ayat references. Fourth, it evaluated context relevance only and did not assess answer faithfulness, answer relevance, citation accuracy, retrieval recall, or expert interpretation. Accordingly, this study presents the evaluation as an initial proof of concept rather than as a definitive benchmark.

8.6. Implications for Further Evaluation

The findings identify several requirements for a controlled ThemePath-RAG benchmark. Future experiments should use the same language representation for the query, thematic path, and candidate ayat, or use a multilingual embedding model that maps all representations into a shared semantic space. Vector RAG and ThemePath-RAG should also use the same Qur'anic translation, final top-*k* context budget, and evaluation procedure.

A stronger benchmark should compare at least four retrieval conditions:

1. Vector RAG using dense verse-level retrieval;
2. BM25 RAG using lexical verse-level retrieval;
3. thematic-path retrieval without query-aware evidence scoring; and

4. ThemePath-RAG with weighted lexical, semantic, and path-based scoring followed by global top- k pruning.

In addition to RAGAS-based context relevance, future evaluation should include gold-ayah Precision@ k , Recall@ k , MRR, nDCG@ k , citation accuracy, answer faithfulness, answer relevance, context efficiency, and expert assessment. For Qur'anic QA, an additional domain-specific metric should measure correct correct surah-ayah citation and correct distinction between Qur'anic text, tafsir, and hadith.

Therefore, the present experiment should be regarded as the first empirical step toward ThemePath-RAG evaluation. It establishes the feasibility of operationalizing curated thematic paths, identifies the importance of calibrated evidence scoring and global pruning, and provides a concrete basis for a future controlled benchmark.

9. Promising Cross-Domain Applications of ThemePath-RAG Beyond Qur'anic Question Answering

The bibliometric and critical analyses suggest that the main opportunity for ThemePath-RAG is not limited to religious question answering. The broader problem appears whenever a corpus contains expert-authored thematic organization and canonical evidence units. In this section, we expand the motivating case from Qur'anic QA to several promising domains as listed in Table 8: library classification and digital libraries, Indonesian legal codes, medical guidelines, educational taxonomies, and policy or government document systems. These domains differ in content, risk level, and evaluation criteria, but they share the same structural retrieval problem: a retrieved category or endpoint is often too broad to be used directly as final evidence.

Table 8. Promising Domains for ThemePath-RAG Beyond Qur'anic Question Answering.

Domain	Curated Path Example	Canonical Evidence Unit	Pruning Criteria	Primary Risk
Library classification	Knowledge domain → class → subject heading → collection	Bibliographic record, abstract, subject heading, full-text passage	Subject overlap, meta-data proximity, source type, publication context	Overbroad recommendation
Indonesian legal codes	Legal area → law → chapter → article → clause	Article, paragraph, clause, explanatory note, SOP step	Article specificity, jurisdiction, clause match, authority level	Wrong legal citation or advice
Medical guidelines	Disease category → diagnosis → treatment → contraindication	Guideline statement, evidence grade, clinical note	Patient constraints, guideline section, evidence grade, contraindication match	Unsafe medical suggestion
Education and curriculum	Competency → topic → subtopic → learning outcome	Learning outcome, module section, assessment item	Grade level, prerequisite, Bloom level, topic match	Mismatched difficulty or objective
Policy and compliance	Policy area → regulation → requirement → procedure	Regulation clause, audit item, required document	Actor, obligation type, effective date, procedure match	Misstated obligation

9.1. Case 1: Library Classification and Digital Library Retrieval

Library and information science provides a strong domain for ThemePath-RAG because library collections are already organized through classification schedules, subject headings, controlled vocabularies, thesauri, collection hierarchies, and bibliographic metadata. A digital library item is rarely just a text chunk. It is connected to class numbers, subject descriptors, authors, publication types, collections, languages, and institutional

repositories. Recent RAG studies in digital libraries and cultural heritage show the potential of retrieval-augmented systems for reading promotion, cultural heritage access, and collection-oriented search [168,169]. However, standard RAG pipelines often flatten bibliographic records, abstracts, full-text documents, or metadata into independent chunks.

A library classification path can be treated as a curated thematic path. For instance, a user query about “machine learning for digital library recommendation” may map to a path such as Computer Science → Artificial Intelligence → Machine Learning → Recommender Systems → Digital Libraries. Endpoint-only retrieval may return every document under recommender systems or digital libraries, but only a subset may address the user’s specific need. ThemePath-RAG can retrieve the classification path, expand candidate bibliographic records and full-text passages under the path, and prune evidence using query-to-record relevance, subject-heading overlap, metadata proximity, citation availability, and source type priority. This is particularly useful for scholarly recommendation, reference advisory systems, literature mapping, and OPAC-style conversational search.

In this domain, the canonical evidence unit is not necessarily a verse or legal article. It may be a bibliographic record, abstract, paragraph, controlled subject heading, classification notation, or cited work. The citation requirement is also different. The system should preserve bibliographic metadata, distinguish item-level evidence from collection-level summaries, and avoid recommending documents solely because they share a broad class. Therefore, ThemePath-RAG can serve as a bridge between knowledge organization systems and LLM-based retrieval.

9.2. Case 2: Indonesian Legal Codes and Regulatory Question Answering

Legal corpora are highly suitable for ThemePath-RAG because they are naturally hierarchical and citation-sensitive. Indonesian legal materials may include constitutions, statutes, government regulations, ministerial regulations, regional regulations, articles, paragraphs, clauses, explanatory notes, and court decisions. Legal QA has already become an active RAG application area, including legal document access, contract law assistance, compliance review, public procurement law, legal knowledge graphs, and Indonesian legal QA [46,118,120,123,170–173]. Indonesian RAG studies also include government document chatbots, Dukcapil SOP question answering, and Indonesian legal document processing [173–175]. More specifically, Indonesian civil law has already been modeled as a graph-based RAG corpus using 2,128 articles from the KUH Perdata, with Neo4j representing the legal hierarchy across books, chapters, sections, and articles [8]. The benchmark evaluated multiple LLMs using RAGAS metrics, including faithfulness, answer relevancy, and context entity recall. These findings provide direct empirical motivation for ThemePath-RAG: graph-structured legal retrieval improves contextual grounding, but additional query-aware pruning is needed when a retrieved legal path or node contains multiple potentially relevant provisions.

A legal question rarely asks for a broad topic alone. It usually requires a precise article, clause, exception, condition, or procedural step. For example, a user may ask about requirements for a civil registration procedure, criminal liability, contractual validity, land administration, or public-service obligation. A thematic path may look like Civil Administration → Population Registration → Identity Document → Required Procedure. If the system retrieves all provisions under the endpoint, it may include irrelevant articles, definitions, transitional rules, sanctions, or administrative procedures. ThemePath-RAG can reduce this problem by using the legal hierarchy as a retrieval route and then pruning evidence at the article, clause, or SOP-step level.

For Indonesian legal codes, the scoring function should be adapted. Source priority may rank statutory text above explanatory notes, secondary commentary, and generated

summaries. Citation specificity should reward exact references to law number, year, article, paragraph, and clause. Graph proximity may capture whether a clause belongs to the same chapter or subsection as the retrieved legal path. Lexical scoring can be useful because legal terms, article numbers, and institutional names often require exact matching. The output should distinguish binding legal text from explanatory interpretation and should include a caution that the system is not a substitute for professional legal advice.

9.3. Case 3: Medical Guidelines and Clinical Pathways

Medical RAG is one of the largest domain-specific clusters in the bibliometric analysis. Recent studies include hybrid medical RAG, knowledge-graph medical QA, clinical decision support, electronic health records, and evidence-based hallucination reduction [28,32,52,61,63,90,107,109]. Medical corpora often contain curated structures such as disease categories, diagnostic criteria, treatment pathways, contraindications, dosage instructions, evidence grades, and clinical workflow stages. This makes them a natural candidate for ThemePath-RAG.

A medical query such as “What is the recommended next step for an elderly patient with suspected condition X and contraindication Y?” cannot be answered by retrieving a broad disease page alone. It requires movement along a clinical path, for example, Disease Category → Diagnosis → Risk Stratification → Treatment Option → Contraindication. Endpoint-only retrieval may return many guideline statements, but only some apply to the specific patient scenario. ThemePath-RAG can retrieve the guideline path and prune evidence using query constraints, patient context, evidence grade, guideline section, and contraindication matching.

The medical case also demonstrates why evidence pruning is not only an efficiency mechanism. It is a safety mechanism. Broad retrieval may expose the LLM to irrelevant treatments or outdated context. Query-aware pruning helps ensure that generated answers cite the most specific guideline statements and do not mix diagnosis, treatment, and patient-education content without clear separation.

9.4. Case 4: Educational Taxonomies and Curriculum-Based QA

Educational corpora are often organized by curriculum standards, competencies, learning outcomes, topics, subtopics, Bloom’s taxonomy levels, assessment rubrics, and learning resources. RAG has been applied to educational policy understanding, adaptive tutoring, educational knowledge graphs, automatic question generation, and higher education support [44,62,124–127,129,130,176]. These applications often require retrieval that respects curricular structure rather than merely retrieving semantically similar paragraphs.

A student question such as “Can you explain recursion for a beginner and give me practice questions at the application level?” implies a path such as Programming → Functions → Recursion → Beginner Explanation → Bloom Application. A standard vector retriever may return advanced material, examples from unrelated languages, or assessment items with mismatched difficulty. ThemePath-RAG can retrieve the curriculum path, expand candidate learning materials, and prune them using topic match, learning outcome alignment, grade level, prerequisite structure, and assessment level. In this case, citation specificity may refer not only to sources but also to curriculum identifiers, course learning outcomes, or module sections.

9.5. Case 5: Policy, Government, and Compliance Documents

Policy and compliance documents are another promising domain because they combine hierarchy, authority, and procedural specificity. RAG studies have addressed regulatory compliance, audit policy, pharmacy policy assistance, public procurement, and government-document chatbots [114–116,119,121,122,174]. These corpora often have a

structured path such as Policy Area → Regulation → Requirement → Procedure → Evidence Document. A query may require a specific compliance obligation rather than all documents under a policy theme.

ThemePath-RAG can exploit this structure by using policy paths as retrieval routes and pruning candidate evidence according to obligation type, jurisdiction, effective date, responsible actor, procedural step, and required documentation. This is especially relevant for public-sector chatbots, university accreditation systems, audit assistants, and regulatory compliance copilots. The system should preserve source hierarchy and avoid presenting policy summaries as binding obligations unless the underlying regulation is cited.

10. Evaluation Agenda and Benchmark Design

10.1. Research Questions

A future controlled benchmark should address the following questions. First, under matched language representations, evidence budgets, and gold evidence annotations, under which conditions can ThemePath-RAG improve retrieval quality relative to flat Vector RAG, BM25 RAG, hybrid retrieval, and graph-based retrieval? Second, can query-aware evidence pruning reduce context redundancy while preserving evidence coverage and answer completeness? Third, can ThemePath-RAG improve citation accuracy in canonical text question answering? Fourth, how should lexical similarity, semantic similarity, path similarity, graph proximity, source priority, and citation specificity be calibrated using an independent validation set? Fifth, can ThemePath-RAG generalize beyond Qur'anic QA to other curated thematic corpora?

10.2. Dataset Construction

For Qur'anic QA, a benchmark dataset should include factual questions, thematic questions, multi-hop questions, comparative questions, historical-context questions, tafsir-oriented questions, and hadith-related questions. Each question should be annotated with gold ayah references, optional tafsir evidence, optional hadith evidence, relevant thematic path, and expected answer type. A controlled Qur'anic benchmark should also ensure that the query, thematic-path labels, candidate ayat, and retrieval baseline use the same language representation or a verified multilingual representation. The present proof of concept used Indonesian representations for Vector RAG and English representations for ThemePath-RAG; therefore, future evaluation should eliminate this confound. Gold ayah references should be annotated independently to support Precision@*k*, Recall@*k*, MRR, and nDCG@*k* evaluation. For cross-domain validation, similar benchmark structures can be created for library classification, Indonesian legal codes, medical guidelines, and educational curricula. A library benchmark may annotate gold bibliographic records, subject headings, classification paths, and recommended evidence passages. A legal benchmark may annotate law numbers, articles, paragraphs, clauses, explanatory notes, and procedural steps. A medical benchmark may annotate guideline sections, evidence grades, contraindications, and clinical pathway nodes. An educational benchmark may annotate learning outcomes, prerequisite topics, grade levels, and assessment levels. This cross-domain design would test whether ThemePath-RAG is a general retrieval framework rather than a single-domain solution.

10.3. Baselines

The present proof of concept compares Vector RAG with the weighted ThemePath-RAG implementation only. The additional baselines below are proposed for a future controlled benchmark. Minimum baselines should include BM25-only RAG, vector-only RAG, hybrid BM25 plus vector RAG, graph endpoint retrieval with all connected evidence,

graph neighborhood retrieval, thematic path retrieval without pruning, and ThemePath-RAG. Stronger baselines may include HippoRAG-inspired PPR retrieval and PathRAG-inspired path pruning.

10.4. Metrics

The present proof-of-concept evaluation reports RAGAS-based context relevance as a retrieval-stage diagnostic metric. A future benchmark should extend this assessment with gold-evidence retrieval metrics, evidence-efficiency metrics, answer-quality metrics, citation-quality metrics, and domain-specific safety measures. As shown in Table 9, recommended retrieval metrics include Precision@k, Recall@k, MRR, nDCG@k, gold evidence hit rate, evidence redundancy ratio, and context reduction ratio. Recommended answer metrics include faithfulness, groundedness, citation accuracy, answer relevance, completeness, hallucination rate, and source distinction accuracy. Domain-specific metrics should reflect the canonical evidence units of each corpus. For Qur’anic QA, additional metrics include correct surah and ayah citation; correct distinction between Qur’an, tafsir, and hadith; no unsupported religious ruling; evidence sufficiency; and interpretive caution. For legal QA, metrics should include article or clause citation accuracy and authority-level distinction. For library systems, metrics should include subject-heading alignment, bibliographic metadata correctness, and recommendation relevance. For medical QA, metrics should include guideline-section accuracy, evidence-grade alignment, and unsafe-recommendation rate. For education, metrics should include learning-outcome alignment, grade-level match, and assessment-level match.

Table 9. Current and Recommended Evaluation Dimensions for ThemePath-RAG.

Dimension	Example Metrics
Current PoC metric	RAGAS context relevance
Retrieval quality	Precision@k, Recall@k, MRR, nDCG@k, gold evidence hit rate
Evidence efficiency	Context length, context reduction ratio, evidence redundancy ratio
Answer grounding	Faithfulness, groundedness, hallucination rate
Citation quality	Citation accuracy, source distinction accuracy
Domain sensitivity	Evidence sufficiency, unsupported-claim rate, interpretive caution

11. Discussion

11.1. Why ThemePath-RAG Is Not Just Hybrid Retrieval

ThemePath-RAG does not merely combine vector retrieval and lexical scoring. Its novelty lies in retrieval architecture. It defines thematic paths as retrieval objects, treats endpoint themes as candidate spaces rather than final evidence, and introduces query-aware pruning over canonical units. Vector retrieval can be used to find thematic paths, and BM25 can be used to prune candidate evidence, but the framework is defined by the staged use of curated paths and canonical evidence units.

The proof-of-concept result should not be interpreted as evidence that this architecture is empirically superior to Vector RAG. Under the present evaluation setting, Vector RAG achieved higher context relevance. The contribution of ThemePath-RAG is therefore its explicit retrieval architecture for exploiting curated semantic paths and globally pruning canonical evidence, whose effectiveness requires further controlled evaluation.

11.2. Relation to HippoRAG and PathRAG

This paper provides a conceptual comparison with HippoRAG and PathRAG, but it does not include a direct empirical benchmark against either method. Such comparisons remain necessary before claims about relative effectiveness can be made. HippoRAG is valuable for associative retrieval across constructed knowledge graphs. It uses PPR to

retrieve related passages through graph activation. ThemePath-RAG adopts the insight that retrieval should not stop at isolated chunks, but it differs by starting from curated thematic paths rather than OpenIE-generated graph nodes. PathRAG is valuable because it identifies redundancy as a central weakness of graph-based RAG and proposes path-level pruning. ThemePath-RAG adopts the anti-redundancy principle but differs by using curated thematic paths as semantic routes and pruning canonical evidence under those paths.

11.3. Generalization Potential

The expanded case studies indicate that ThemePath-RAG is best understood as a framework for corpora whose structure has already been organized by experts. Its generalization depends on three conditions: the corpus must contain meaningful paths, the terminal or intermediate nodes must connect to citeable evidence units, and the domain must benefit from pruning evidence before generation. Qur'anic QA, library classification, legal codes, medical guidelines, curriculum standards, and compliance documents satisfy these conditions in different ways. These domains appear structurally compatible with the ThemePath-RAG formulation because they contain curated paths and citeable evidence units. However, empirical generalization beyond Qur'anic QA has not yet been established. The proposed cross-domain applications should therefore be interpreted as motivated design hypotheses for future evaluation rather than validated deployments.

11.4. Practical Efficiency

ThemePath-RAG may offer practical efficiency benefits because thematic-path retrieval can narrow the candidate evidence space before final evidence scoring. In static corpora, thematic paths and associated evidence metadata can be precomputed, allowing lexical and semantic scoring to operate over a restricted candidate set rather than the entire corpus.

However, the present proof-of-concept evaluation does not measure retrieval latency, computational cost, memory consumption, or end-to-end token efficiency. Although the final evidence budget was closely matched between the two systems, ThemePath-RAG returned slightly longer contexts on average than Vector RAG (103.13 versus 77.96 words). Therefore, no empirical efficiency advantage is claimed. Future work should report retrieval time, scoring time, context length, token usage, and answer-generation cost under matched hardware and model settings.

12. Limitations and Open Challenges

This paper combines a critical survey, bibliometric analysis, framework proposal, and an initial proof-of-concept evaluation in Qur'anic question answering. The experiment demonstrates the operational feasibility of ThemePath-RAG, but it does not constitute a comprehensive benchmark. Under the present paired evaluation, Vector RAG achieved significantly higher RAGAS context relevance than ThemePath-RAG. The result should not be interpreted as evidence of superiority for ThemePath-RAG across retrieval architectures or domains.

Several factors limit causal interpretation. Although both systems returned approximately three final ayat, Vector RAG used Indonesian question and ayat representations, whereas ThemePath-RAG used English translations. The proof of concept also used heuristic similarity weights rather than weights selected through an independent validation procedure. In addition, the evaluation used a single LLM-based RAGAS evaluator and did not include manually annotated gold ayah references, human expert assessment, answer-faithfulness evaluation, or citation-accuracy evaluation. Future work should conduct a language-matched, multi-baseline benchmark with gold evidence labels and expert review.

Another limitation is that curated thematic structures may vary in quality across domains. Some corpora have expert-authored taxonomies, while others require manual or semi-automatic construction. ThemePath-RAG also faces challenges in multi-path queries, evidence diversity, interpretive ambiguity, cross-lingual retrieval, and expert evaluation cost.

In sacred, legal, and medical domains, evaluation should involve domain experts. Citation accuracy and groundedness cannot always be judged reliably by automatic metrics alone. Future studies should also explore how to represent uncertainty, disagreement, and multiple valid interpretations in the final answer.

13. Conclusions

RAG research has rapidly advanced from flat vector retrieval to graph-based, agentic, multimodal, and memory-inspired systems. However, most existing methods still assume that knowledge is either stored as flat chunks or must be automatically structured into an indexing graph. This overlooks a class of high-value corpora that already possess curated thematic structures.

This paper identified this gap through a critical survey supported by bibliometric evidence from 2815 Scopus-indexed RAG-related records. The bibliometric analysis shows that RAG research is rapidly expanding and globally active, but its dominant vocabulary remains centered on general LLM-retrieval integration, information retrieval, question answering, semantics, and knowledge graphs. It does not reveal a mature cluster around curated thematic paths or canonical evidence pruning.

To address this gap, this paper proposed ThemePath-RAG as a retrieval framework for curated thematic corpora and operationalized it in a Qur'anic question-answering proof of concept. ThemePath-RAG treats curated thematic paths as high-recall semantic routes, expands candidate canonical evidence, and applies query-aware scoring and global pruning to produce a compact final evidence set.

The paired proof-of-concept evaluation showed that ThemePath-RAG can be implemented with top- $m = 3$ thematic-path selection and global top- $k = 3$ ayat pruning. However, Vector RAG achieved higher RAGAS context relevance than ThemePath-RAG under the current configuration (0.920 versus 0.798; $p < 0.001$). Thus, the experiment establishes feasibility and identifies evidence-selection challenges, rather than demonstrating superiority over conventional vector retrieval.

The broader contribution of ThemePath-RAG is to frame curated thematic organization as a retrieval resource that can guide evidence selection before generation. Its potential application to legal codes, clinical guidelines, educational curricula, policy documents, and library taxonomies remains a hypothesis requiring cross-domain validation. Future work should test this framework using language-matched benchmarks, stronger retrieval baselines, gold evidence annotations, answer-level metrics, citation evaluation, and expert assessment.

Author Contributions: Conceptualization, W.M., D.A.D., A.H.N. and A.O.; methodology, W.M. and A.H.N.; software, W.M. and A.H.N.; validation, W.M., D.A.D., A.H.N. and A.O.; formal analysis, W.M., D.A.D. and A.H.N.; investigation, W.M. and A.H.N.; resources, W.M., A.H.N. and Y.M.; data curation, W.M. and A.H.N.; writing—original draft preparation, W.M.; writing—review and editing, W.M., D.A.D. and A.H.N.; visualization, W.M.; supervision, D.A.D., A.H.N., A.O. and Y.M.; funding acquisition, W.M. and D.A.D. All authors have read and agreed to the published version of the manuscript.

Funding: The APC is covered by INTI International University.

Data Availability Statement: The materials supporting the ThemePath-RAG proof-of-concept experiment are publicly available in the ThemePathRAG repository at <https://github.com/WindaMo/ThemePathRAG> (accessed on 21 June 2026). The repository contains the 150 Indonesian Qur’anic QA questions and their corresponding English query set; ThemePath-RAG retrieval outputs; RAGAS context-relevance results; retrieved contexts; selected thematic paths; candidate ayat; lexical, semantic, and path-similarity values; combined evidence scores; and the experiment notebook and execution script used for the thematic-path retrieval, evidence-scoring, and RAGAS evaluation workflow. The repository does not redistribute the raw Scopus metadata export or the complete upstream Qur’anic thematic corpus. The thematic organization and source metadata are derived from the *al-Qur’an al-Hadi* thematic index; users should consult the original source for the full underlying corpus and its terms of use.

Acknowledgments: During the preparation of this work, the author(s) used OpenAI ChatGPT (GPT-5.5 Thinking model) for language editing and refinement. After using this tool, the author(s) reviewed, revised, and edited the content as needed, taking full responsibility for the content of the published article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.t.; Rocktäschel, T.; et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. In Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Red Hook, NY, USA, Online, 6–12 December 2020.
- Karpukhin, V.; Oguz, B.; Min, S.; Lewis, P.; Wu, L.; Edunov, S.; Chen, D.; Yih, W.T. Dense Passage Retrieval for Open-Domain Question Answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Online; Webber, B., Cohn, T., He, Y., Liu, Y., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, November 2020; pp. 6769–6781. [\[CrossRef\]](#)
- Izacard, G.; Caron, M.; Hosseini, L.; Riedel, S.; Bojanowski, P.; Joulin, A.; Grave, E. Unsupervised Dense Information Retrieval with Contrastive Learning. *arXiv* **2022**, arXiv:2112.09118.
- Santhanam, K.; Khattab, O.; Saad-Falcon, J.; Potts, C.; Zaharia, M. ColBERTv2: Effective and Efficient Retrieval via Lightweight Late Interaction. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Seattle, WA, USA; Carpuat, M., de Marneffe, M.C., Meza Ruiz, I.V., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, July 2022; pp. 3715–3734. [\[CrossRef\]](#)
- Edge, D.; Trinh, H.; Cheng, N.; Bradley, J.; Chao, A.; Mody, A.; Truitt, S.; Metropolitansky, D.; Ness, R.O.; Larson, J. From Local to Global: A Graph RAG Approach to Query-Focused Summarization. *arXiv* **2025**, arXiv:2404.16130.
- Guo, Z.; Xia, L.; Yu, Y.; Ao, T.; Huang, C. LightRAG: Simple and Fast Retrieval-Augmented Generation. In *Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2025, Suzhou, China*; Christodoulopoulos, C., Chakraborty, T., Rose, C., Peng, V., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 4–9 November 2025; pp. 10746–10761. [\[CrossRef\]](#)
- Pan, S.; Luo, L.; Wang, Y.; Chen, C.; Wang, J.; Wu, X. Unifying Large Language Models and Knowledge Graphs: A Roadmap. *IEEE Trans. Knowl. Data Eng.* **2024**, *36*, 3580–3599. [\[CrossRef\]](#)
- Nabila, D.; Nasution, A.H.; Murakami, Y.; Koos, S.; Ergun, A.E. Modeling and Benchmarking GraphRAG for Indonesian Legal Question Answering. *Artif. Intell. Lang. Model.* **2026**, *1*, 1–12. [\[CrossRef\]](#)
- Chen, B.; Guo, Z.; Yang, Z.; Chen, Y.; Chen, J.; Liu, Z.; Shi, C.; Yang, C. PathRAG: Pruning Graph-based Retrieval Augmented Generation with Relational Paths. In Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI), Singapore, 20–27 January 2026; Volume 40, pp. 30183–30191. [\[CrossRef\]](#)
- Gutiérrez, B.J.; Shu, Y.; Gu, Y.; Yasunaga, M.; Su, Y. HippoRAG: Neurobiologically Inspired Long-Term Memory for Large Language Models. In *Proceedings of the Advances in Neural Information Processing Systems, 10–15 December 2024*; Globerson, A., Mackey, L., Belgrave, D., Fan, A., Paquet, U., Tomczak, J., Zhang, C., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2024; Volume 37, pp. 59532–59569. [\[CrossRef\]](#)
- Haveliwalla, T.H. Topic-sensitive PageRank. In Proceedings of the 11th International Conference on World Wide Web, New York, NY, USA, 7–11 May 2002; WWW ’02; pp. 517–526. [\[CrossRef\]](#)
- Gutiérrez, B.J.; Shu, Y.; Qi, W.; Zhou, S.; Su, Y. From RAG to Memory: Non-Parametric Continual Learning for Large Language Models. In *Proceedings of the 42nd International Conference on Machine Learning. PMLR, 13–19 July 2025*; Proceedings of Machine Learning Research: New York, NY, USA, 2025; Volume 267, pp. 21497–21515.

13. Khalila, Z.; Nasution, A.H.; Monika, W.; Onan, A.; Murakami, Y.; Radi, Y.B.I.; Osmani, N.M. Investigating Retrieval-Augmented Generation in Quranic Studies: A Study of 13 Open-Source Large Language Models. *Int. J. Adv. Comput. Sci. Appl.* **2025**, *16*, 1361–1371. [CrossRef]
14. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef] [PubMed]
15. Elsevier. Scopus: A Comprehensive Abstract and Citation Database for Impact Makers. 2026. Available online: <https://www.elsevier.com/products/scopus> (accessed on 13 June 2026).
16. Baas, J.; Schotten, M.; Plume, A.; Côté, G.; Karimi, R. Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quant. Sci. Stud.* **2020**, *1*, 377–386. [CrossRef]
17. Pereira, V.; Basilio, M.P.; Santos, C.H.T. PyBibX—A Python library for bibliometric and scientometric analysis powered with artificial intelligence tools. *Data Technol. Appl.* **2025**, *59*, 302–337. [CrossRef]
18. Saeed, T.; Wang, B. Large Language Models for Generative Recommendation: A Systematic Review of Data-Centric Taxonomy, Evaluation, and Human-Centric Analytics. *Int. J. Data Sci. Anal.* **2026**, *22*, 130. [CrossRef]
19. Billah, S.M.; Yusof, R.J.B.R. Exploring the Role of RAG in Cloud-Based Monolithic Chatbots: A Systematic Review. In Proceedings of the 2025 IEEE 23rd Student Conference on Research and Development (SCORED), Kuala Lumpur, Malaysia, 25–26 November 2025; pp. 1–6. [CrossRef]
20. Tsallis, C.; Papageorgas, P.; Munteanu, R.A.; Dellagi, S. Large Language and Foundation Models for Machinery Health Monitoring: A Systematic Review. *Appl. Sci.* **2026**, *16*, 2493. [CrossRef]
21. Choi, M.; Ahsan, I.; Yu, H.; Choe, T.; Kim, M. The Semantic Design Space of Retrieval-Augmented Recommender Systems: A Systematic Review of LLM-Based Approaches. *Comput. Mater. Contin.* **2026**, *88*, 3. [CrossRef]
22. Fatawi, I.; Bilad, M.; Asy'ari, M. The role of prompt engineering in enhancing LLMs: A systematic review of applications and ethical implications. *IAES Int. J. Artif. Intell. (IJ-AI)* **2026**, *15*, 1071–1086. [CrossRef]
23. Trad, F.; Yammine, R.; Charafeddine, J.; Chakhtoura, M.; Rahme, M.; El-Hajj Fuleihan, G.; Chehab, A. Streamlining Systematic Reviews with Large Language Models Using Prompt Engineering and Retrieval Augmented Generation. *BMC Med. Res. Methodol.* **2025**, *25*, 130. [CrossRef] [PubMed]
24. Shah, J.; Gade, S.R.; Ponna, D.; Patel, K.A. Cloud-Native AI and Generative AI on AWS: A Systematic Review and a Proposed Unified Model. In Proceedings of the 2026 7th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), Goathgaun, Morang, Nepal, 7–9 January 2026; pp. 1710–1717. [CrossRef]
25. Mikulić, I.; Vlaić, M.; Delač, G.; Šilić, M.; Vladimir, K. Integrating External Knowledge with LLMs: A Systematic Review of RAG Approaches. In Proceedings of the 2025 MIPRO 48th ICT and Electronics Convention, Opatija, Croatia, 2–6 June 2025; pp. 93–98. [CrossRef]
26. Smajić, A.; Karlović, R.; Bobanović Dasko, M.; Lorencin, I. Large Language Models for Structured and Semi-Structured Data, Recommender Systems and Knowledge Base Engineering: A Survey of Recent Techniques and Architectures. *Electronics* **2025**, *14*, 3153. [CrossRef]
27. Murugaraj, K.; Lamsiyah, S.; Theobald, M. ExpandFuse: A Hybrid Retrieval Framework with Query Expansion and Topic-Aware Reranking for Multi-Hop Question Answering. In Proceedings of the 2025 IEEE International Conference on Big Data (BigData), Macau, China, 8–11 December 2025; pp. 722–729. [CrossRef]
28. Umadevi, K.S.; Rajeshwari, S.R.; Sai Likhitha, P. Medical RAG: Hybrid Retrieval & Re-Ranking. In Proceedings of the 2025 International Conference on Responsible, Generative and Explainable AI (ResGenXAI), Bhubaneswar, India, 10–12 September 2025; pp. 1–6. [CrossRef]
29. Sun, Y. Adaptive Hybrid Retrieval-Augmented Generation with Context-Aware Confidence Control. In Proceedings of the 2025 4th International Conference on Electronic Information Technology (EIT), Chengdu, China, 22–24 August 2025; pp. 891–895. [CrossRef]
30. Abirami, S.; Joshma, M.J.; Jayavarthini, A.; Joshi, A.; Gajendran, M.K. Evaluating Lexical, Dense, and Hybrid Retrieval Pipelines for RAG. In Proceedings of the 2025 IEEE Pune Section International Conference (PuneCon), Pune, India, 12–14 December 2025; pp. 1–6. [CrossRef]
31. Hu, C.; Huang, Y.; Kuang, J.; Dai, B.; Peng, Y.; Xiao, Y.; Su, Y. Mitigating Hallucinations in Discipline Inspection QA: A Two-Stage RAG Framework with Late Interaction and Reranking. *Electronics* **2026**, *15*, 541. [CrossRef]
32. Aljohani, B.; Alsanoosy, T. Enhancing Medical Question Answering with LLMs via a Hybrid Retrieval-Augmented Generation Framework. *Information* **2026**, *17*, 133. [CrossRef]
33. Ahmad, S.; Nezami, Z.; Hafeez, M.; Raza Zaidi, S.A. Benchmarking Vector, Graph and Hybrid Retrieval Augmented Generation (RAG) Pipelines for Open Radio Access Networks (ORAN). In Proceedings of the 2025 IEEE 36th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), İstanbul, Türkiye, 1–4 September 2025; pp. 1–6. [CrossRef]

34. Achyar, M.K.U.; Widyawan; Pratama, A.R. RAG Chatbot Architecture for Law & Crime News Using Hybrid Retrieval and Small Language Model. In Proceedings of the 2025 5th International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA), Yogyakarta, Indonesia, 17–19 December 2025; pp. 588–593. [\[CrossRef\]](#)
35. Altınok, H.; Tekgoz, H.; Erdoğan, H.; Uz, H. A Comparative Analysis of Retrieval-Augmented Generation Architectures with Semantic Hashing for Enterprise Knowledge Systems. In Proceedings of the 2026 5th International Informatics and Software Engineering Conference (IISEC), Ankara, Türkiye, 5–6 February 2026; pp. 343–347. [\[CrossRef\]](#)
36. Hamza, R.M.; Ajam, G.E. Context-Aware Intelligent Data Integration Approach: A practical Hybrid Retrieval Agent. In Proceedings of the 2026 2nd International Conference on Computing and Emerging Sciences (ICCES '26), Erbil, Iraq, 4–5 February 2026; pp. 121–127. [\[CrossRef\]](#)
37. Li, R.; Mao, S.; Zhu, C.; Yang, Y.; Tan, C.; Li, L.; Mu, X.; Liu, H.; Yang, Y. Enhancing Pulmonary Disease Prediction Using Large Language Models With Feature Summarization and Hybrid Retrieval-Augmented Generation: Multicenter Methodological Study Based on Radiology Report. *J. Med. Internet Res.* **2025**, *27*, e72638. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Shang, Y.; Ke, Z.; Lin, P.; Ren, Q.; Zhang, W.; Wang, X.; Li, X.; Gong, F.; Wang, S.; Wang, B.; et al. Empowering knowledge graphs with hybrid retrieval-augmented generation for the intelligent mix scheme of mass concrete. *Case Stud. Constr. Mater.* **2025**, *23*, e04979. [\[CrossRef\]](#)
39. Abdolinejad, F.; Eftekhari, M. Augmenting RAG with Nonnegative Matrix Factorization-Driven Semantic Chunking in Embedding Space. *J. Supercomput.* **2026**, *82*, 224. [\[CrossRef\]](#)
40. Lavarec, E.; Du, Y. Evaluating document chunking approaches for retrieval augmented generation in editorial content. *IAES Int. J. Artif. Intell. (IJ-AI)* **2026**, *15*, 1909–1918. [\[CrossRef\]](#)
41. Koay, X.K.; Ong, L.Y.; Goh, P.Y. Structure-Aware Chunking for Complex Tables in Retrieval-Augmented Generation Systems. *Emerg. Sci. J.* **2026**, *10*, 184–205. [\[CrossRef\]](#)
42. Zhang, L.; Ning, Y. Improving construction contract question answering through embedding optimization and semantic chunking in large language models. *Adv. Eng. Inform.* **2026**, *69*, 104027. [\[CrossRef\]](#)
43. Moreno-Cediel, A.; Garcia-Lopez, E.; Garcia-Cabot, A.; De-Fitero-Dominguez, D. Optimising retrieval performance in RAG systems: A new growing window semantic chunking strategy to address weak semantic boundaries. *Knowl.-Based Syst.* **2026**, *331*, 114896. [\[CrossRef\]](#)
44. Niu, Y.; Rong, X. Improving Retrieval-Augmented Generation for Educational Policy Understanding via Structure-Aware Text Chunking. In Proceedings of the 2026 3rd International Conference on Informatics Education and Computer Technology Applications (IECA 2026), Shanghai, China, 16–18 January 2026; pp. 1093–1100. [\[CrossRef\]](#)
45. Li, X.; Xue, T. HTR-GEN: A Structure-Aware Framework for Hierarchical Table Retrieval and Generation. In Proceedings of the 2025 5th International Conference on Computer Systems (ICCS), Xi'an, China, 26–28 September 2025; pp. 77–81. [\[CrossRef\]](#)
46. Lee, S.; Kim, N.; Lee, J. Structural Chunking: A Semantic-Structural Integrated Method for Retrieval-Augmented Generation. In Proceedings of the 2026 International Conference on Electronics, Information, and Communication (ICEIC), Macau, China, 18–21 January 2026; pp. 1–6. [\[CrossRef\]](#)
47. Jaiswal, S.; Bisht, P.; Kansara, K.; Datta, M.S. Comparison of Chunking Techniques Across Diverse Document Types in NLP Retrieval Tasks. In Proceedings of the 2025 International Conference on Responsible, Generative and Explainable AI (ResGenXAI), Bhubaneswar, India, 10–12 September 2025; pp. 1–6. [\[CrossRef\]](#)
48. Widiiswa, R.A.N.; Henry, M.M.; Imani, A.; Pardamean, B. Retrieval-Augmented Generation Chunking Strategies for Financial Document Analysis: A Systematic Literature Review. In Proceedings of the 2026 30th International Conference on Information Technology (IT), Žabljak, Montenegro, 24–28 February 2026; pp. 1–4. [\[CrossRef\]](#)
49. Qu, L.; Zhao, X.; Zhang, C.; Li, G. GraphRAG-Vet: A Knowledge Graph-Augmented Large Language Model for Precision Bovine Disease Diagnosis. *Computers* **2026**, *15*, 203. [\[CrossRef\]](#)
50. Wang, X.; Fang, J. Claim Knowledge Graph Construction and GraphRAG-Based Question-Answering System. *Buildings* **2026**, *16*, 845. [\[CrossRef\]](#)
51. Li, M.; Qin, R. DualGraphRAG: A Dual-View Graph-Enhanced Retrieval-Augmented Generation Framework for Reliable and Efficient Question Answering. *Appl. Sci.* **2026**, *16*, 2221. [\[CrossRef\]](#)
52. Chen, S.; Chen, T. Co-MedGraphRAG: A Collaborative Large–Small Model Medical Question-Answering Framework Enhanced by Knowledge Graph Reasoning. *Information* **2026**, *17*, 247. [\[CrossRef\]](#)
53. Jiang, B.; Liu, Z.; Wang, N.; Li, Z.; Shi, Y.; Lin, B. Process-Oriented Dual-Layer Knowledge GraphRAG for Reservoir Engineering Decision Support. *Processes* **2025**, *13*, 3230. [\[CrossRef\]](#)
54. Knollmeyer, S.; Caymazer, O.; Grossmann, D. Document GraphRAG: Knowledge Graph Enhanced Retrieval Augmented Generation for Document Question Answering Within the Manufacturing Domain. *Electronics* **2025**, *14*, 2102. [\[CrossRef\]](#)
55. Wang, C.; Fu, Y.; Wang, C.; Wang, L.; Zheng, Q.; Zhang, H.; Li, M.; Meng, F. CRAG-IKG: A Conflict- and Reliability-Aware GraphRAG Framework for Noisy Industrial Knowledge Graphs. In Proceedings of the 2025 5th International Conference on

- Electronic Communication, Computer Science and Technology (ECCST), Qinhuangdao, China, 26–28 December 2025; pp. 373–377. [\[CrossRef\]](#)
56. Lv, X.; Feng, Y.; Zheng, J. SL-MERK: Synthetic Lethality Mechanism Explainer based on GraphRAG and Knowledge Graph. In Proceedings of the 2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Copenhagen, Denmark, 14–18 July 2025; pp. 1–6. [\[CrossRef\]](#) [\[PubMed\]](#)
 57. Xu, J.; Zhang, H.; Zhang, H.; Lu, J.; Xiao, G. ChatTf: A Knowledge Graph-Enhanced Intelligent Q&A System for Mitigating Factuality Hallucinations in Traditional Folklore. *IEEE Access* **2024**, *12*, 162638–162650. [\[CrossRef\]](#)
 58. Utami, L.; Rachmi, H.; Hidayatulloh, S. A Hybrid TF-IDF and Knowledge Graph-Enhanced Retrieval-Augmented Generation Framework with Large Language Models for Domain-Aware Question Answering. *J. Appl. Data Sci.* **2026**, *7*, 866–881. [\[CrossRef\]](#)
 59. Lin, S.; Shao, S.; Liu, X.; Su, H. FusionGraphRAG: An Adaptive Retrieval-Augmented Generation Framework for Complex Disease Management in the Elderly. *Information* **2026**, *17*, 138. [\[CrossRef\]](#)
 60. Jovanovski, D.; Stojcheva, M.; Dodevska, M.; Lameski, P.; Mishkovski, I.; Gjorgjevikj, D. An Empirical Study of Knowledge Graph-Enhanced RAG for Information Security Compliance. *Information* **2026**, *17*, 389. [\[CrossRef\]](#)
 61. Lecu, A.; Groza, A.; Hawizy, L. Reducing Hallucinations in Medical AI: A Knowledge Graph-Augmented Retrieval System for Evidence-Based Age-Related Macular Degeneration Information. *IEEE Access* **2025**, *13*, 210624–210639. [\[CrossRef\]](#)
 62. Dong, C.; Yuan, Y.; Chen, K.; Cheng, S.; Wen, C. How to Build an Adaptive AI Tutor for Any Course Using Knowledge Graph-Enhanced Retrieval-Augmented Generation (KG-RAG). In Proceedings of the 2025 14th International Conference on Educational and Information Technology (ICEIT), Guangzhou, China, 14–16 March 2025; pp. 152–157. [\[CrossRef\]](#)
 63. Yadav, V.; Gaurav; Rana, A.; Sharma, S. MedRAG-Agent: Medical Query Resolution By Employing A Multi-Agent, Knowledge Graph-Enhanced RAG-Based AI Framework. In Proceedings of the 2025 IEEE 6th Global Conference for Advancement in Technology (GCAT), Bangalore, India, 24–26 October 2025; pp. 1–4. [\[CrossRef\]](#)
 64. Chen, W.; Zhou, Y.; Yan, J.; He, Y.; Lu, H.; Tian, Z. APT-ArgusQA: Knowledge Graph-Enhanced Large Language Model-based Question Answering for Advanced Persistent Threats. In Proceedings of the 2025 IEEE 6th International Conference on Computer, Big Data, Artificial Intelligence (ICCBD+AI), Xiamen, China, 21–23 November 2025; pp. 1–6. [\[CrossRef\]](#)
 65. Flüh, M.; Kim, S.Y.; Schneider, C.V.; Geisler, S. FAIR GraphRAG: A Retrieval-Augmented Generation Approach for Semantic Data Analysis. In Proceedings of the 2025 IEEE International Conference on Knowledge Graph (ICKG), Limassol, Cyprus, 13–14 November 2025; pp. 90–97. [\[CrossRef\]](#)
 66. Huang, S.; Cheng, J. StructReason: A Multi-hop Retrieval-Augmented Generation Framework via Iterative Reasoning Chains and PCST Structural Refinement. In Proceedings of the 2025 5th International Conference on Communication Technology and Information Technology (ICCTIT), Guangzhou, China, 26–28 December 2025; pp. 201–206. [\[CrossRef\]](#)
 67. Ding, L.; Ding, N.; Tao, Q.; Shi, P. Enhancing graph multi-hop reasoning for question answering with LLMs: An approach based on adaptive path generation. *J. Intell. Inf. Syst.* **2025**, *63*, 1455–1485. [\[CrossRef\]](#)
 68. Jian, Y.; Lan, L.; Luo, Z. CogNav: Cognitive Navigation for Complex Multi-Hop Retrieval via Chain-of-Thought Reasoning. In Proceedings of the 2nd International Conference on Digital Society, Information Science and Risk Management (ICDIR 2026), Shenyang, China, 23–25 January 2026; pp. 114–117. [\[CrossRef\]](#)
 69. Huang, H. H2RAG: A Hub-Aware Hypergraph Retrieval-Augmented Generation Framework for Multi-Hop Reasoning. In Proceedings of the 2026 5th International Conference on Electronics Technology and Artificial Intelligence (ETAI), Harbin, China, 6–8 March 2026; pp. 362–365. [\[CrossRef\]](#)
 70. Huang, Y.; Yang, L.; Yang, X.H.; Xu, X. Retrieval-Augmented Generation for Multi-Hop Question Answering Based on Structured Planning. *ACM Trans. Knowl. Discov. Data* **2026**, *20*, 44:1–44:20. [\[CrossRef\]](#)
 71. Zhang, X.; Zhao, F.; Liu, Y.; Chen, P.; Wang, Y.; Wang, X.; Ma, D.; Xu, H.; Chen, M.; Li, H. TreeQA: Enhanced LLM-RAG with logic tree reasoning for reliable and interpretable multi-hop question answering. *Knowl.-Based Syst.* **2025**, *330*, 114526. [\[CrossRef\]](#)
 72. Wang, H.; Wang, T.; Sun, Z.; Li, H.; Cao, Z.; Feng, L.; Wang, D. A Query-Driven Graph Retrieval Framework with Adaptive Pruning for Multi-Hop Question Answering. *Electronics* **2026**, *15*, 1263. [\[CrossRef\]](#)
 73. Wang, S.; Wang, Z.; Qu, C.; Yin, Z. Multi-level retrieval with representation alignment enhances cross-document evidence synthesis for scientific knowledge generation. *Appl. Soft Comput.* **2026**, *191*, 114649. [\[CrossRef\]](#)
 74. Zai, X.; Tan, X.; Wang, X.; Liu, Q.; Xu, X.; Zhang, W. PRoH: Dynamic Planning and Reasoning over Knowledge Hypergraphs for Retrieval-Augmented Generation. In Proceedings of the ACM Web Conference 2026, Dubai, United Arab Emirates, 13–17 April 2026; pp. 4256–4267. [\[CrossRef\]](#)
 75. Hamid, M.R.A.; El-Regaily, S.A.; Aref, M.M. Can Large Language Models Perform Retrieval-Augmented Generation as Multi-Hop Reasoning Over Knowledge Graphs? In Proceedings of the 2025 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), Cairo, Egypt, 17–18 September 2025; pp. 360–365. [\[CrossRef\]](#)
 76. Wang, J.; Shen, H.; Xie, B.; Chen, Y.; Zhao, W.; Wang, X.; Hong, Y.; Fu, C.; Pan, Z.; Sun, L.; et al. Agentic Graph-RAG: A Multi-Agent Framework for Robust, Decomposed Multi-Hop Reasoning. In Proceedings of the 2025 11th International Conference on Computer and Communications (ICCC), Chengdu, China, 12–15 December 2025; pp. 574–578. [\[CrossRef\]](#)

77. Fu, R.; Wang, Y.; Xu, T.; Liu, Y.; Tang, W.; Wu, W.; Ma, X.; Fong, S. S-Path-RAG: Semantic-Aware Shortest-Path Retrieval Augmented Generation for Multi-Hop Knowledge Graph Question Answering. In Proceedings of the ACM Web Conference 2026, Dubai, United Arab Emirates, 29 June–3 July 2026; pp. 4057–4068. [\[CrossRef\]](#)
78. Kamlakshya, T. Agentic RAG Systems for Real-Time Financial Decision Making: A Multi-Agent Framework with Model Context Protocol Integration. In Proceedings of the 2025 IEEE 2nd International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS), Bangalore, India, 29–30 August 2025; pp. 1–3. [\[CrossRef\]](#)
79. Hariharan, M.; Barma, S.; Arvapalli, S.; Sheela, E. Agentic RAG for Software Testing with Hybrid Vector-Graph and Multi-Agent Orchestration. In Proceedings of the 2025 IEEE International Conference on Data and Software Engineering (ICoDSE), Batam, Indonesia, 28–29 October 2025; pp. 54–59. [\[CrossRef\]](#)
80. Hajaghaie, A.; Thulasiram, R.K. Agentic Portfolio Construction: A Multi-Agent Architecture for LLM-Driven Financial Asset Allocation. In Proceedings of the 2025 3rd International Conference on Foundation and Large Language Models (FLLM), Vienna, Austria, 25–28 November 2025; pp. 192–201. [\[CrossRef\]](#)
81. Liang, C.; Cui, Y.; Shi, R.; Zha, G.; Yin, X.; Xiao, M.; Xu, D.; Duan, X.; Huang, B. GeoAgentic-RAG: A Multi-Agent framework for autonomous geospatial reasoning and visual insight generation with LLM. *Int. J. Appl. Earth Obs. Geoinf.* **2026**, *147*, 105195. [\[CrossRef\]](#)
82. Tanasă, A.M.; Oprea, S.V.; Băra, A. Designing an Architecture of a Multi-Agentic AI-Powered Virtual Assistant Using LLMs and RAG for a Medical Clinic. *Electronics* **2026**, *15*, 334. [\[CrossRef\]](#)
83. Chondamrongkul, N.; Kyaw, M.P.P.; Ko, S.M.; Paing, P.P.; Than Swe, M.K.; Hongthong, T. RepoAI: Automated code refactoring through multi-agent LLM orchestration and retrieval-augmented generation. *Sci. Comput. Program.* **2026**, *253*, 103477. [\[CrossRef\]](#)
84. Roumeliotis, K.I.; Sapkota, R.; Karkee, M.; Tselikas, N.D. Agentic AI With Orchestrator-Agent Trust: A Modular Visual Classification Framework With Trust-Aware Orchestration and RAG-Based Reasoning. *IEEE Access* **2026**, *14*, 26965–26982. [\[CrossRef\]](#)
85. Ahi, K.; Hsieh, C.H.; Fenger, G. LLMs and LVMs for agentic AI: A GPU-accelerated multimodal system architecture for RAG-grounded, explainable, and adaptive intelligence. In *Proceedings of the Photomask Technology 2025*, Monterey, CA, USA, 22–25 September 2025; Melvin, L.S., III, Philipsen, V., Eds.; Proceedings of SPIE; International Society for Optics and Photonics: Bellingham, WA, USA, 2025; Volume 13687, p. 136871R. [\[CrossRef\]](#)
86. Hadee, A.N.A.; Riznee, M. Code Prism: A Multi-Agent, Multi-LLM, Semantic Indexing Artifact for Regulatory Code Audits—A Design Science Research Study. In Proceedings of the 2025 1st International Conference on Emerging Innovation and Digital Technology (ICEIDT), Malé, Maldives, 24–25 November 2025; pp. 23–27. [\[CrossRef\]](#)
87. Salehi, S.; Singh, Y.; Horst, K.K.; Hathaway, Q.A.; Erickson, B.J. Agentic AI and Large Language Models in Radiology: Opportunities and Hallucination Challenges. *Bioengineering* **2025**, *12*, 1303. [\[CrossRef\]](#) [\[PubMed\]](#)
88. Liu, Z.; Kou, J.; Zhang, W.; Gu, C.; Fang, X.; Huang, Z.; Yuan, H.; Li, H.; Lu, X.; Yin, A.; et al. Comprehensive Evaluation of AI Hallucination and Novel UV-Oriented Framework toward Safe and Trustworthy AI. In Proceedings of the 2024 7th International Conference on Universal Village (UV), Boston, MA, USA, 19–22 October 2024; pp. 1–136. [\[CrossRef\]](#)
89. Gu, C.; Zhang, W.; Huang, Z.; Kou, J.; Liu, Z.; Zhao, C.; Liu, C.; Zhang, L.; Lin, W.; Wang, Z.; et al. LENS: Layers of Evaluation of Hallucination in GenAI Systems. In Proceedings of the 2024 7th International Conference on Universal Village (UV), Boston, MA, USA, 19–22 October 2024; pp. 1–85. [\[CrossRef\]](#)
90. Pawlik, L.; Deniziak, S. Reducing Hallucinations in Medical AI Through Citation Enforced Prompting in RAG Systems. *Appl. Sci.* **2026**, *16*, 3013. [\[CrossRef\]](#)
91. Pai, K.C.; Hsu, W.C. Fine-tuning small language models for industrial retrieval-augmented generation: Efficiency, factuality, and hallucination analysis. *Comput. Stand. Interfaces* **2026**, *98*, 104163. [\[CrossRef\]](#)
92. Tamilselvi, P.; Ahmed Amin, K.; Vigneshvaran, K.; Rohith, S. Benchmarking Quantized LLMS for Faithfulness in Resource-Constrained RAG. In Proceedings of the 2026 Second International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI), Erode, India, 2–4 March 2026; pp. 762–765. [\[CrossRef\]](#)
93. Krainovskikh, V.; Samigulin, T. Domain-Adapted Retrieval-Augmented Generation for Technical Documentation: Enhancing Reliability and Faithfulness in Technical QA. *IEEE Access* **2026**, *14*, 36016–36024. [\[CrossRef\]](#)
94. Olariu, M.E.; Buinceanu, V.G.; Simionescu, C.; Dospinescu, O.; Georgescu, R.; Tudor, C.; Iftene, A.; Bores, A.M. RO-FIN-LLM: A Benchmark with LLM-as-a-Judge and Human Evaluators for Romanian Tax and Accounting. *Systems* **2026**, *14*, 244. [\[CrossRef\]](#)
95. Papageorgiou, G.; Sarlis, V.; Maragoudakis, M.; Magnisalis, I.; Tjortjis, C. Evaluating Faithfulness in Agentic RAG Systems for e-Governance Applications Using LLM-Based Judging Frameworks. *Big Data Cogn. Comput.* **2025**, *9*, 309. [\[CrossRef\]](#)
96. Wang, Y.; Zhang, Y.; Xu, K.; Xu, Y.; Wang, Y.; Feng, Q. CORB-RAG: A Comprehensive Evaluation Benchmark for Retrieval-Augmented Generation Systems in the Chinese Telecommunications Operator Domain. In Proceedings of the 2025 IEEE 5th International Conference on Computer Communication and Artificial Intelligence (CCAI), Haikou, China, 23–25 May 2025; pp. 503–508. [\[CrossRef\]](#)

97. More, R. A Unified Evaluation Framework for Grounded LLM Architectures: Comparative Analysis of RAG, Self-RAG, and Agentic RAG. In Proceedings of the 2025 5th International Conference on Artificial Intelligence and Signal Processing (AISP), Vijayawada, India, 22–24 November 2025; pp. 1–5. [\[CrossRef\]](#)
98. Nishisako, S.; Higashi, T.; Wakao, F. Reducing Hallucinations and Trade-Offs in Responses in Generative AI Chatbots for Cancer Information: Development and Evaluation Study. *JMIR Cancer* **2025**, *11*, e70176. [\[CrossRef\]](#) [\[PubMed\]](#)
99. Wysocka, M.; Wysocki, O.; Delmas, M.; Mutel, V.; Freitas, A. Large Language Models, scientific knowledge and factuality: A framework to streamline human expert evaluation. *J. Biomed. Inform.* **2024**, *158*, 104724. [\[CrossRef\]](#) [\[PubMed\]](#)
100. Zhang, H.; Li, D.; Nie, X. Mitigating Execution Hallucinations and Computational Inflation in Agentic RAG via Strict Protocol Boundaries. *Electronics* **2026**, *15*, 1805. [\[CrossRef\]](#)
101. Wallat, J.; Heuss, M.; Rijke, M.d.; Anand, A. Correctness is not Faithfulness in Retrieval Augmented Generation Attributions. In Proceedings of the 2025 International ACM SIGIR Conference on Innovative Concepts and Theories in Information Retrieval (ICTIR), Padua, Italy, 18 July 2025; Zamani, H., Dietz, L., Piwowarski, B., Bruch, S., Eds.; ACM: New York, NY, USA, 2025; pp. 22–32. [\[CrossRef\]](#)
102. Noguera, A.; Mogollón-Benavides, A.L.; Niño-Mojica, M.D.; Rua, S.; Sanin-Villa, D.; Tejada, J.C. Applications and Challenges of Retrieval-Augmented Generation (RAG) in Maternal Health: A Multi-Axial Review of the State of the Art in Biomedical QA with LLMs. *Sci* **2025**, *7*, 148. [\[CrossRef\]](#)
103. Shukla, D.; Shirote, S.; Soma, O.; Sorte, A.; Banchhor, S.; Takale, D. MediSense: AI-Based Dual Summarization of Clinical Reports for Healthcare Professionals and Patients. In Proceedings of the 2025 3rd DMIHER International Conference on Artificial Intelligence in Healthcare, Education and Industry (IDICAIHEI), Wardha, India, 28–29 November 2025; pp. 1–6. [\[CrossRef\]](#)
104. Feng, Y.; Wang, J.; He, R.; Zhou, L.; Li, Y. A retrieval-augmented knowledge mining method with deep thinking LLMs for biomedical research and clinical support. *GigaScience* **2025**, *14*, g1af109. [\[CrossRef\]](#) [\[PubMed\]](#)
105. Tang, W.; Chen, R.; Long, X.; Yu, D.; Zhao, S.; Chen, B. Medical large language models and systems in the clinical application of spinal diseases: Current status, challenges, and future prospects. *J. Orthop. Transl.* **2026**, *57*, 101050. [\[CrossRef\]](#) [\[PubMed\]](#)
106. Rahulprasath, S.; Pranav Harshan, S.; Kabilash, P.V.; Lakshithraj, A.; Sreemathy, J. AI in Healthcare: Simplifying Medical Reports for Enhanced Patient Comprehension. In Proceedings of the 2025 International Conference on Emerging Technologies in Computing and Communication (ETCC), Bangalore, India, 26–27 June 2025; pp. 1–6. [\[CrossRef\]](#)
107. Thio, S.; Lewis, M.; Denaxas, S.; Dobson, R.J.B. Unlocking electronic health records: A hybrid graph RAG approach to safe clinical AI for patient QA. *Front. Digit. Health* **2026**, *8*. [\[CrossRef\]](#) [\[PubMed\]](#)
108. Kulshreshtha, A.; Choudhary, A.; Taneja, T.; Verma, S. Enhancing Healthcare Accessibility: A RAG- Based Medical Chatbot Using Transformer Models. In Proceedings of the 2024 International Conference on IT Innovation and Knowledge Discovery (ITIKD), Manama, Bahrain, 13–15 April 2025; pp. 1–4. [\[CrossRef\]](#)
109. Raputri, E.; Teguh, A.J.; Hidayah, S.N.; Anom, A.K.; Setiawan, F.; Qomariyah, N.N. Retrieval-Augmented LLMs with Indonesian Clinical Trials Guidelines: A Comparative Study. In Proceedings of the 2025 8th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Yogyakarta, Indonesia, 11 December 2025; pp. 459–464. [\[CrossRef\]](#)
110. Li, W.; Zhang, Y.; Wang, C.; Li, Y.; He, X.; Wang, A.L.; Xu, M.; Zhang, F.; Sun, H.; Wang, K.; et al. CARE: A clinical agentic reasoning engine to enhance real-World diagnostic accuracy via structured medical reasoning. *Expert Syst. Appl.* **2026**, *313*, 131476. [\[CrossRef\]](#)
111. Wang, J.F.; Chang, C.C.; Chiang, T.M.; Yeh, T.C.; Cheng, E.; Lee, Y.T.; Chen, H.I. An innovative X-RAG technique combined with GPT-4o for summarizing medical information from EHR and EMR to assist doctors in clinical decision-making effectively and efficiently. *Health Inform. J.* **2025**, *31*, 1–16. [\[CrossRef\]](#) [\[PubMed\]](#)
112. Guiducci, L.; Saulle, C.; Dimitri, G.M.; Valli, B.; Alpini, S.; Tenti, C.; Rizzo, A. Dialogical AI for Cognitive Bias Mitigation in Medical Diagnosis. *Appl. Sci.* **2026**, *16*, 710. [\[CrossRef\]](#)
113. Berkowitz, J.S.; Srinivasan, A.; Acitores Cortina, J.M.; Fatapour, Y.; Tatonetti, N.P. Biomedical text normalization through generative modeling. *J. Biomed. Inform.* **2025**, *167*, 104850. [\[CrossRef\]](#) [\[PubMed\]](#)
114. Kalyanasundaram, T.; Bussari, S.; Sundaram, D.; Sharma, K.K. Engineering Reliable Retrieval-Augmented Generation for Regulatory and Compliance Systems. In Proceedings of the SoutheastCon 2026, Huntsville, AL, USA, 13–15 March 2026; pp. 1–5. [\[CrossRef\]](#)
115. Bharucha, D.; Palanisamy, N.; Reddy, M. ReguQuery: An Agentic Framework For Automated Regulatory Compliance Analysis Using Retrieval-Augmented Generation. In Proceedings of the 2026 6th Biennial International Conference on Nascent Technologies in Engineering (ICNTE), Navi Mumbai, India, 16–17 January 2026; pp. 1–6. [\[CrossRef\]](#)
116. Arshad, U.; Corsar, D.; Nkisi-Orji, I. Integrating KGs and ontologies with RAG for personalised summarisation in regulatory compliance. In Proceedings of the SICSA REALLM Workshop 2024, Aberdeen, UK, 17 October 2024; CEUR Workshop Proceedings; Volume 3822; pp. 56–61.

117. Xu, Y.; Xue, C.; Zhu, T.; Tian, Y.; Hu, J.N.; Tong, L. Research on the Technical Framework of Legal Compliance Review of Major Decision-Making based on Large Language Model. In Proceedings of the 2025 IEEE 5th International Conference on Applied Mathematics, Modeling and Computer Simulation (AMMCS), Wuhan, China, 23–24 August 2025; pp. 1–6. [\[CrossRef\]](#)
118. Zhai, H. Law GraphRAG: An Advanced Legal Question-Answering System. In Proceedings of the 2025 5th International Conference on Artificial Intelligence and Industrial Technology Applications (AIITA), Xi'an, China, 28–30 March 2025; pp. 1407–1410. [\[CrossRef\]](#)
119. Kim, J.; Hur, M.; Min, M. From RAG to QA-RAG: Integrating Generative AI for Pharmaceutical Regulatory Compliance Process. In Proceedings of the 40th ACM/SIGAPP Symposium on Applied Computing, Catania, Italy, 31 March 2025–4 April 2025; pp. 1293–1295. [\[CrossRef\]](#)
120. Yao, S.; Ke, Q.; Wang, Q.; Li, K.; Hu, J. Lawyer GPT: A Legal Large Language Model with Enhanced Domain Knowledge and Reasoning Capabilities. In Proceedings of the 2024 3rd International Symposium on Robotics, Artificial Intelligence and Information Engineering, Singapore, 5–7 July 2024; pp. 108–112. [\[CrossRef\]](#)
121. Kalyanasundaram, T.; Jadav, V. A Compliance-Focused Retrieval-Augmented AI System for Pharmacy Policy Assistance. In Proceedings of the 2025 International Conference on Computer and Applications (ICCA), Bahrain, Bahrain, 22–24 December 2025; pp. 1–6. [\[CrossRef\]](#)
122. Huang, S.; Sun, C.; Ning, M.; Yang, Y.; Ma, C.; Zhong, J.; Shu, K.; Shi, F.; Khajepour, A. DriveLegal: Toward legally compliant driving via trustworthy hybrid retrieval-augmented LLMs. *Expert Syst. Appl.* **2026**, *314*, 131593. [\[CrossRef\]](#)
123. Amazou, Y.; Tayalati, F.; Mensouri, H.; Azmani, A.; Azmani, M. Accurate AI Assistance in Contract Law Using Retrieval-Augmented Generation to Advance Legal Technology. *Int. J. Adv. Comput. Sci. Appl.* **2025**, *16*, 1141–1150. [\[CrossRef\]](#)
124. Xu, B.; Tong, R.J.; Li, Y.; Chen, P.; Li, H.; Liang, J.; Fan, X.; Tong, J. An Architectural Framework for Educational Knowledge Graphs (IEEE P2807.6): Ontology Design, Llm Integration, and Adaptive Learning Applications. In Proceedings of the 2025 IEEE Conference on Artificial Intelligence (CAI), Santa Clara, CA, USA, 5–7 May 2025; pp. 1610–1616. [\[CrossRef\]](#)
125. Ho, T.L.; Lam, T.P. EduMSRA: A Multi-Source Educational Research Agent Integrating Retrieval-Augmented Generation and Model Context Protocol for Adaptive Intelligent Tutoring Systems. *Appl. Sci.* **2026**, *16*, 4400. [\[CrossRef\]](#)
126. Golla, F. Enhancing Student Engagement Through AI-Powered Educational Chatbots: A Retrieval-Augmented Generation Approach. In Proceedings of the 2024 21st International Conference on Information Technology Based Higher Education and Training (ITHET), Paris, France, 6–8 November 2024; pp. 1–6. [\[CrossRef\]](#)
127. Maity, S.; Deroy, A.; Sarkar, S. Leveraging In-Context Learning and Retrieval-Augmented Generation for Automatic Question Generation in Educational Domains. In Proceedings of the 16th Annual Meeting of the Forum for Information Retrieval Evaluation, Gandhinagar, India, 12–15 December 2024; pp. 40–47. [\[CrossRef\]](#)
128. Sahana, S.; Ghosh, P.; Anutariya, C. Reimagining Education: An Architectural Blueprint for a Synergistic AI-Driven Learning Environment. In Proceedings of the 2025 IEEE 16th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Berkeley, CA, USA, 29–31 October 2025; pp. 0123–0130. [\[CrossRef\]](#)
129. Németh, R.; Tátrai, A.; Szabó, M.; Zaletnyik, P.T.; Tamási, Á. Exploring the use of retrieval-augmented generation models in higher education: A pilot study on artificial intelligence-based tutoring. *Soc. Sci. Humanit. Open* **2025**, *12*, 101751. [\[CrossRef\]](#)
130. Khan, A.; Chandekar, S.; Baikar, A.; Kanade, A.; Kadam, S.; Kadam, V. Educational Application of RAG-based LLMs. In Proceedings of the 2025 5th International Conference on Evolutionary Computing and Mobile Sustainable Networks (ICECMSN), Coimbatore, India, 24–26 November 2025; pp. 720–725. [\[CrossRef\]](#)
131. Tawfik, M.K.; Ali, M.O.; Mohamed, S.K.; Ezz Elregal, F.S.; Edward, P.E.; Elsayaf, M.; Khattab, D. Mufakkir: RAG-Based Arabic Educational Chatbot for University Students. In Proceedings of the 2025 Twelfth International Conference on Intelligent Computing and Information Systems (ICICIS), Cairo, Egypt, 25–27 November 2025; pp. 706–712. [\[CrossRef\]](#)
132. Tan, A.; Dorneich, M.C.; Cotos, E. Aligning Pedagogy with Generative AI: An Approach to Customizing Educational GPTs. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*; SAGE Publications: Los Angeles, CA, USA, 2025; Volume 69, pp. 894–899. [\[CrossRef\]](#)
133. Hu, W.; Gong, R.; Wu, S.; Li, X. A conversational agent based on contingent teaching model to support collaborative learning activities: Impacts on students' learning performance, self-efficacy and perceptions. *Educ. Technol. Res. Dev.* **2025**, *73*, 3341–3372. [\[CrossRef\]](#)
134. Aljohani, B.; Aljuhani, A. Pioneering agentic retrieval-augmented generation in software quality: A novel framework for code smell detection dynamic retrieval. *PeerJ Comput. Sci.* **2026**, *12*, e3642. [\[CrossRef\]](#)
135. İçöz, B.; Biricik, G. Context-Aware Code Review Automation: A Retrieval-Augmented Approach. *Appl. Sci.* **2026**, *16*, 1875. [\[CrossRef\]](#)
136. Sharanarathi, T.; Polineni, S. Multi-Agent LLM Collaboration for Adaptive Code Review, Debugging, and Security Analysis. In Proceedings of the 2025 International Conference on Mechatronics, Robotics, and Artificial Intelligence (MRAI), Jinan, China, 19–21 June 2025; pp. 541–546. [\[CrossRef\]](#)

137. Elezi, A.; Cico, B.; Hyseni, D. Tuning DeepSeek-Coder-V2-Lite-Base for C# Code Smell Detection: Advancing Towards Task Versatility in Software Maintenance. In Proceedings of the 2025 14th Mediterranean Conference on Embedded Computing (MECO), Budva, Montenegro, 10–14 June 2025; pp. 1–4. [\[CrossRef\]](#)
138. Liao, D.; Pan, S.; Sun, X.; Ren, X.; Huang, Q.; Xing, Z.; Jin, H.; Li, Q. A³A3-CodGen: A Repository-Level Code Generation Framework for Code Reuse With Local-Aware, Global-Aware, and Third-Party-Library-Aware. *IEEE Trans. Softw. Eng.* **2024**, *50*, 3369–3384. [\[CrossRef\]](#)
139. Jaoua, I.; Sghaier, O.B.; Sahraoui, H. Combining Large Language Models with Static Analyzers for Code Review Generation. In Proceedings of the 2025 IEEE/ACM 22nd International Conference on Mining Software Repositories (MSR), Ottawa, ON, Canada, 28–29 April 2025; pp. 174–186. [\[CrossRef\]](#)
140. Yang, L.; Chen, Y. Integrating RAG and LLM for Automated Code Review in Practice. In Proceedings of the 2025 10th International Conference on Computer and Information Processing Technology (ISCIPT), Fushun, China, 12–14 September 2025; pp. 591–596. [\[CrossRef\]](#)
141. Denuri, T.T.L.; Ragukanthan, S.; Wijesekara, W.A.D.N.; Prageeth, M.D.; Silva, A.P.; Attanayaka, L. Archelon AI: Programming Assistant for Legacy Codebases. In Proceedings of the 2025 10th International Conference on Information Technology Research (ICITR), Colombo, Sri Lanka, 8–11 December 2025; pp. 1–6. [\[CrossRef\]](#)
142. Jiang, C.; Zhang, P.; Ni, Y.; Wang, X.; Peng, H.; Liu, S.; Fei, M.; He, Y.; Xiao, Y.; Huang, J.; et al. Multimodal retrieval-augmented generation for financial documents: Image-centric analysis of charts and tables with large language models. *Vis. Comput.* **2025**, *41*, 7657–7670. [\[CrossRef\]](#)
143. Loo, C.W.; Qian Leong, Z.; Ong, J.K.; Thong Chong, H.; Yu, Y.P.; Ming Lim, T. Retrieval-Augmented Generation with GPT-4o-mini: Integrating Configurable Chunking, Hybrid Search, and Multimodal Image Retrieval. In Proceedings of the 2025 IEEE International Conference on Computation, Big-Data and Engineering (ICCBE), Penang, Malaysia, 27–29 June 2025; pp. 862–866. [\[CrossRef\]](#)
144. Chang, C.Y.; Khanum, A.; Su, S.G.; Tsai, M.H.; Hsu, H.H.; Chen, W.T.; Lu, H.W. Optimizing Construction Safety: Multimodal Prompts For Automated Image Descriptions In Daily Construction Reports. In Proceedings of the 2025 IEEE International Conference on Image Processing Workshops (ICIPW), Anchorage, AK, USA, 14–17 September 2025; pp. 263–268. [\[CrossRef\]](#)
145. Zhang, C.; Lin, K.; Yang, Z.; Wang, J.; Li, L.; Lin, C.C.; Liu, Z.; Wang, L. MM-Narrator: Narrating Long-form Videos with Multimodal In-Context Learning. In Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024; pp. 13647–13657. [\[CrossRef\]](#)
146. Ni, T.; Yuan, X.; Li, S.; Ni, W. Privacy-Preserving Multimodal Reasoning for Internet of Things: A Retrieval-Augmented Large Language and Vision Assistant Framework. *IEEE Internet Things Mag.* **2026**, *9*, 113–122. [\[CrossRef\]](#)
147. Suram, K.; A J, H.; Kurian, A.; Augustine, K.; Majeed A, R. Vision-Language Transformers for Medical Report Synthesis: A Multimodal Approach. In Proceedings of the 2025 IEEE 4th International Conference for Advancement in Technology (ICONAT), Goa, India, 19–21 September 2025; pp. 1–5. [\[CrossRef\]](#)
148. Markin, E.I.; Zuparova, V.V.; Martyshkin, A.I. Integration of Large Language Models and Computer Vision Algorithms in LMS: A Methodology for Automated Verification of Software Tasks and Multimodal Analysis of Educational Data. In Proceedings of the 2025 International Russian Smart Industry Conference (SmartIndustryCon), Sochi, Russian Federation, 24–28 March 2025; pp. 777–781. [\[CrossRef\]](#)
149. Miao, J.; Lu, D.; Wang, Z. A Multimodal RAG Framework for Housing Damage Assessment: Collaborative Optimization of Image Encoding and Policy Vector Retrieval. In Proceedings of the 2025 International Conference on Artificial Intelligence and Product Design, New York, NY, USA, 18–20 July 2025; AIPD '25; pp. 281–285. [\[CrossRef\]](#)
150. He, H.; Yuan, X.; Wu, K.; Ni, W. Federated Retrieval-Augmented Generation for Cybersecurity in Resource-Constrained IoT and Edge Environments: A Deployment-Oriented Scoping Review. *Electronics* **2026**, *15*, 1409. [\[CrossRef\]](#)
151. Zaw, H.M.; Khant Naing, H.; Myat, A.K.; Paing Linn, H.; Lin, T.; Khaing, T.T. Browser-Embedded, Legal-Aware Cybersecurity Co-Pilot for Myanmar with RAG, Multilingual Defense and Privacy. In Proceedings of the 2025 6th International Conference on Advanced Information Technologies (ICAIT), Yangon, Myanmar, 3 November 2025; pp. 1–6. [\[CrossRef\]](#)
152. Belmahjoub, M.; Benhiba, L. Causal-Aware Knowledge Graph Enhanced RAG for Predictive Cybersecurity Intelligence: A Framework for Attack Progression Analysis and Consequence Prediction. In Proceedings of the 2025 12th International Conference on Soft Computing & Machine Intelligence (ISCMCI), Rio de Janeiro, Brazil, 21–23 November 2025; pp. 136–141. [\[CrossRef\]](#)
153. Gregory, J.; Liao, Q. Autonomous Cyberattack with Security-Augmented Generative Artificial Intelligence. In Proceedings of the 2024 IEEE International Conference on Cyber Security and Resilience (CSR), London, United Kingdom, 2–4 September 2024; pp. 270–275. [\[CrossRef\]](#)
154. Gokcimen, T.; Das, B. A novel system for strengthening security in large language models against hallucination and injection attacks with effective strategies. *Alex. Eng. J.* **2025**, *123*, 71–90. [\[CrossRef\]](#)

155. Pu, X.; Zhang, Y. Cultivating Cybersecurity Talent: Localized RAG Approach Integrating Pattern Recognition and Document Analysis. *Int. J. Pattern Recognit. Artif. Intell.* **2026**, *40*, 2550034. [[CrossRef](#)]
156. Cheng, Z.; Sun, J.; Gao, A.; Quan, Y.; Liu, Z.; Hu, X.; Fang, M. Secure Retrieval-Augmented Generation Against Poisoning Attacks. In Proceedings of the 2025 IEEE International Conference on Big Data (BigData), Macau, China, 8–11 December 2025; pp. 1799–1806. [[CrossRef](#)]
157. Abd Elhakeem, M.G.; Abdullah, A.O.; Elhag, S.W.; Mohamed, E.H. Aqrag: Advanced quranic retrieval-augmented generation for low-resource question answering. *Neural Comput. Appl.* **2026**, *38*, 282. [[CrossRef](#)]
158. Khanduja, N.; Kumar, D.N.; Arun Chauhan, D. A Retrieval-Augmented Generation Model for Faith-Aligned QA in Bhagvat Gita. In Proceedings of the 2025 International Conference on Intelligent and Secure Engineering Solutions (CISES), Greater Noida Gautam Budh Nagar, India, 11–13 August 2025; pp. 1524–1528. [[CrossRef](#)]
159. Obaid, S.; Bawany, N.Z. SeerahGPT: Retrieval Augmented Generation based Large Language Model. In Proceedings of the 2024 18th International Conference on Open Source Systems and Technologies (ICOSST), Lahore, Pakistan, 26–27 December 2024; pp. 1–7. [[CrossRef](#)]
160. Singh, A.; Shrivastava, P.; Misra, R. Chatbots for Religious Punjabi Texts Using Retrieval-Augmented Generation (RAG) for Factual Accuracy: A Review. In Proceedings of the 2025 IEEE 4th International Conference on Technology, Engineering, Management for Societal impact using Marketing, Entrepreneurship and Talent (TEMSMET), New Delhi, India, 8–10 October 2025; pp. 1–6. [[CrossRef](#)]
161. Sun, S. CIRR: Causal-Invariant Retrieval-Augmented Recommendation with Faithful Explanations under Distribution Shift. In Proceedings of the 2025 6th International Conference on Computer Science and Management Technology, Xiamen, China, 26–28 December 2025; pp. 888–894. [[CrossRef](#)]
162. Bara, A.; Oprea, S.V. AI-Augmented Bibliometric Framework: A Paradigm Shift with Agentic AI for Dynamic, Snippet-Based Research Analysis. *arXiv* **2026**, arXiv:2511.21745.
163. Robertson, S.E.; Walker, S. Some Simple Effective Approximations to the 2-Poisson Model for Probabilistic Weighted Retrieval. In *Proceedings of the SIGIR'94: Proceedings of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, Organised by Dublin City University*; Springer: London, UK, 1994; pp. 232–241. [[CrossRef](#)]
164. Trivedi, H.; Balasubramanian, N.; Khot, T.; Sabharwal, A. Interleaving Retrieval with Chain-of-Thought Reasoning for Knowledge-Intensive Multi-Step Questions. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Toronto, Canada; Rogers, A., Boyd-Graber, J., Okazaki, N., Eds.; Association for Computational Linguistics: Stroudsburg, PA, USA, 2023; pp. 10014–10037. [[CrossRef](#)]
165. Bosma, M.; Chi, E.; Ichter, B.; Le, Q.V.; Schuurmans, D.; Wang, X.; Wei, J.; Xia, F.; Zhou, D. Chain-Of-Thought Prompting Elicits Reasoning in Large Language Models. In Proceedings of the Advances in Neural Information Processing Systems 35, New Orleans, LA, USA, 28 November–9 December 2022; pp. 24824–24837. [[CrossRef](#)]
166. Namratha, N.; Naidu, S.S.S.; Sahoo, K.S.; Singla, M.; Praneeth, B.V.S.S.S.R. GeetaVani: A Retrieval-Augmented LLM Framework for Contextual Dialogue from the Bhagavad Geeta. In Proceedings of the 2025 13th International Conference on Intelligent Systems and Embedded Design (ISED), Raipur, India, 17–19 December 2025; pp. 654–661. [[CrossRef](#)]
167. Es, S.; James, J.; Espinosa Anke, L.; Schockaert, S. RAGAs: Automated Evaluation of Retrieval Augmented Generation. In Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations, St. Julians, Malta, 17–22 March 2024; pp. 150–158. [[CrossRef](#)]
168. Li, Z. Equipping a Virtual Reading Promoter for Digital Libraries with Retrieval-Augmented Generation. In Proceedings of the 2025 ACM/IEEE Joint Conference on Digital Libraries (JCDL), Dekalb, IL, USA, 15–19 December 2025; pp. 313–315. [[CrossRef](#)]
169. Kelly, P.; Schild, J.; Jafari, A.H. FolkRAG: A retrieval-augmented generation system for cultural heritage materials. *Neural Comput. Appl.* **2025**, *37*, 20281–20297. [[CrossRef](#)]
170. Júnior, E.F.P.D.S.; Baptista, C.d.S.; Alves, A.L.F.; Mendes, F.I.d.L. Using Retrieval-Augmented Generation to Improve Access to Legal Documents. In Proceedings of the 2026 International Conference on Semantic Computing (ICSC), Laguna Hills, CA, USA, 2–4 February 2026; pp. 328–335. [[CrossRef](#)]
171. Anhein, M.A.G.; Wiharja, K.R.S. Integrating Knowledge Graphs and Semantic Retrieval for Indonesian Legal Question Answering. In Proceedings of the 2025 5th International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA), Yogyakarta, Indonesia, 17–19 December 2025; pp. 389–394. [[CrossRef](#)]
172. Ongiris, J.G.; Darari, F.; Tobing, B.C.; Faisal, D.R.; Lee, O. Benchmarking KG-based RAG Systems: A Case Study of Legal Documents. In Proceedings of the Second International Workshop on Retrieval-Augmented Generation Enabled by Knowledge Graphs, Nara, Japan, 2–6 November, 2025; Volume 4079, pp. 69–82.
173. Wilsen, W.; Dewandaru, A.; Candra, M.Z.C. Generation and Visualization of BPMN from Legal Documents. In Proceedings of the 2025 IEEE International Conference on Data and Software Engineering (ICoDSE), Batam, Indonesia, 28–29 October 2025; pp. 232–237. [[CrossRef](#)]

174. Rumambi, F.; Prasetya, D.D.; Widiyaningtyas, T. Groundedness-Aware Retrieval in Government Document Chatbots: Systematic Literature Review and Semantic Alignment Score (SAS) Formulation. *ITEGAM-JETIA* **2026**, *12*, 242–257. [[CrossRef](#)]
175. Rizki, A.; Panjaitan, G.P.H.; Azizy, F.N.; Purwarianti, A.; Utama, N.P. Retrieval-Augmented Question Answering for Dukcapil SOPs: Synthetic-Context Fine-Tuning and Chunking Design. In Proceedings of the 2026 International Seminar on Intelligent Business and Edge-Computing Research (ISIBER), Jakarta, Indonesia, 26–26 February 2026; pp. 745–750. [[CrossRef](#)]
176. Liu, Z.; Agrawal, P.; Singhal, S.; Madaan, V.; Kumar, M.; Verma, P.K. LPItutor: An LLM based personalized intelligent tutoring system using RAG and prompt engineering. *PeerJ Comput. Sci.* **2025**, *11*, e2991. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.