Journal of Emerging Intelligence and Engineering Technologies

Research Article

A Retrieval-Augmented Generative Artificial Intelligence Framework for Domain-Specific Knowledge Discovery and Reliable Question Answering

Razakova Rayxan Saylaubekovna[1], Girish Singh Bisht[2]

[1]Department of Psychology, Mamun University, Urganch, Uzbekistan

[2]Tulas University, Dehradun

Email: razakova_rayxan@mamunedu.uz; Girish@tulas.edu.in

Abstract

Domain-specific question answering requires both linguistic fluency and evidence-grounded accuracy, yet conventional large language models may produce unsupported claims, omit document provenance, or fail to reflect updated institutional knowledge. This paper proposes a retrieval-augmented generative artificial intelligence framework that integrates document normalization, metadata-aware semantic chunking, hybrid sparse–dense retrieval, cross-encoder reranking, evidence-constrained generation, citation verification, and selective abstention. A modular mathematical formulation defines retrieval fusion, context selection, grounded decoding, and confidence-based response control. The evaluation protocol specifies document-level partitioning, expert-annotated questions, retrieval and answer metrics, latency measurement, and error analysis. To illustrate reporting requirements without fabricating empirical evidence, numerical tables and graphs use explicitly hypothetical example values; these are not measurements of an implemented system. The framework provides a reproducible architecture and a testable set of hypotheses concerning answer correctness, evidence faithfulness, retrieval coverage, and operational efficiency. The contribution is a transparent research design for trustworthy knowledge discovery in organizational, technical, academic, and regulatory collections.

Keywords: Retrieval-augmented generation; domain-specific question answering; hybrid retrieval; large language models; evidence grounding; knowledge discovery; hallucination mitigation; citation verification.

Received: 10 July 2025
Revised: 18 August 2025
Accepted: 12 September 2025
Published: 30 December 2025
DOI: 10.66592/vbn369qx

© 2025 The Author(s). Published by the Journal of Emerging Intelligence and Engineering Technologies. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0), which permits use, distribution, and reproduction in any medium for non-commercial purposes, provided the original work is properly cited. https://creativecommons.org/licenses/by-nc/4.0/

1. Introduction

Artificial intelligence systems increasingly serve as natural-language interfaces to information that was previously accessible only through manual search, specialist databases, and document review. In universities, hospitals, engineering firms, government agencies, and enterprises, valuable knowledge is dispersed across policy manuals, technical specifications, reports, procedural documents, frequently asked questions, and archival records. Users rarely know which file contains a relevant statement, and the terminology used in a question may differ substantially from that used in the source. Consequently, effective domain-specific question answering requires more than generating plausible language: a system must identify authoritative evidence, interpret it in context, and communicate an answer whose claims can be inspected.

Large language models (LLMs) have made substantial progress in language

understanding, summarization, and multi-step reasoning. Their parametric knowledge nevertheless presents practical limitations when applied to narrow domains. Training data may not include confidential documents, newly revised policies, or proprietary terminology. A model may respond confidently to an unanswerable question, conflate similar entities, or cite an irrelevant passage. Even when an answer happens to be correct, the user may be unable to determine whether it came from an approved source. These concerns are especially consequential when answers influence compliance, financial operations, safety, academic decisions, or clinical workflows.

Retrieval-augmented generation (RAG) addresses part of this challenge by pairing a language model with an external collection of searchable evidence. Instead of relying exclusively on memorized facts, the system retrieves relevant passages at inference time and supplies them as context for answer generation. The approach introduced by Lewis and colleagues [1] combines parametric and non-parametric memory and motivates architectures that can update information by refreshing the index rather than retraining an entire model. Yet simply appending search results to a prompt does not guarantee reliability. Poor segmentation, lexical mismatch, irrelevant retrieval, contradictory passages, and insufficient verification can all propagate errors into the final response.

Domain-specific corpora create retrieval problems that are distinct from broad open-domain web search. Specialized abbreviations and codes may carry precise meanings, while the same term can have different interpretations across departments. A policy may contain an exception in a footnote, a technical requirement may be split across tables, and a revised document may supersede an older version. Document-level provenance, access permissions, version metadata, and time validity therefore become first-class components of a reliable knowledge system. A suitable architecture must handle these constraints before generation begins.

Traditional lexical search methods, including probabilistic ranking approaches such as BM25 [2], remain useful for exact terminology, part numbers, and named entities. Dense neural retrieval can capture semantic relationships between differently worded queries and passages [3]. Neither method is universally superior: lexical ranking may miss paraphrases, whereas embeddings may overlook rare identifiers or subtle numerical distinctions. Hybrid retrieval combines complementary evidence sources, and a second-stage reranker can examine query–passage interactions more carefully. The quality of the retrieval stage often determines whether a generator receives the information needed to answer accurately.

Knowledge discovery introduces an additional requirement beyond finding a single sentence. A user may ask for a comparison across multiple reports, an explanation of how a process changed, or a synthesis of several clauses. The system must preserve source boundaries while composing a coherent response. It should distinguish direct quotations, inferential summaries, and unresolved ambiguities. When two authoritative documents conflict, a reliable assistant should surface the disagreement rather than invent a reconciled policy. When no sufficient evidence exists, abstaining or requesting clarification may be preferable to producing a confident guess.

Grounding is also an evaluation problem. Conventional overlap metrics such as exact match and token-level F1 can be informative for short factual questions but may understate the quality of valid paraphrases or fail to detect unsupported additions. Modern RAG evaluation separates retrieval relevance, answer relevance, and faithfulness to retrieved evidence [4]. Other frameworks combine automated judging with small human-labeled samples [5]. An

adequate experimental design must therefore assess not only answer correctness but also the evidence behind each claim, citation accuracy, retrieval coverage, and system responsiveness.

This study proposes a modular domain-aware RAG framework organized around ingestion, indexing, retrieval, evidence selection, grounded generation, and post-generation verification. The architecture supports both sparse and dense search, metadata filtering, reranking, and explicit source citation. Its central design principle is that evidence must be traceable through the entire pipeline: from an authorized source document to an indexed chunk, from a retrieved passage to a generated claim, and from that claim to a verifiable citation. This traceability is intended to make failure diagnosis and human review practical.

The research questions are threefold. First, can hybrid retrieval and reranking improve the availability of relevant evidence compared with either sparse or dense retrieval alone? Second, can evidence-constrained prompting and claim-level verification reduce unsupported statements while preserving useful answer coverage? Third, what trade-offs emerge between retrieval depth, context length, latency, and answer quality? These questions are framed as testable hypotheses, not as already established experimental outcomes of the proposed system.

The paper contributes an integrated technical design, mathematical formulations for the principal stages, an evaluation protocol suitable for specialized document collections, and a clearly labeled illustrative results template. It does not claim to have executed a benchmark or trained a new model. This distinction is important because credible applied AI research must separate architectural proposals, established findings from prior literature, and measured results from a real deployment. The remaining sections examine relevant literature, detail the framework, specify reproducible evaluation procedures, and discuss expected benefits, limitations, and directions for validation.

2. Related Work

Research on information retrieval established the foundation for selecting relevant documents from large collections. Probabilistic ranking models estimate relevance using term frequency, document frequency, and document length, with BM25 becoming a widely used strong lexical baseline [2]. In specialized settings, lexical retrieval is particularly valuable when exact phrases, codes, and rare technical expressions matter. However, exact matching may be insufficient when users employ colloquial questions or paraphrase official terminology. This gap motivated neural representation learning for information retrieval, where semantically related queries and passages can be mapped into a shared vector space.

Dense Passage Retrieval (DPR) demonstrated how a dual-encoder architecture could learn representations for questions and passages and perform efficient approximate nearest-neighbor retrieval [3]. The approach separates query encoding from document encoding, enabling precomputation of corpus embeddings. Dense retrieval is appealing for domain-specific collections because it can recover conceptually related text without exact word overlap. Its limitations include sensitivity to domain shift, embedding model choice, and rare tokens. For a corpus containing numerical thresholds, legal clauses, or specialized product identifiers, dense similarity alone may rank a generally relevant passage above the precise authoritative answer.

Lewis et al. [1] introduced retrieval-augmented generation as a combination of neural generation and non-parametric memory. Their formulations show how

retrieved passages can condition an output sequence and how different passages may contribute to generation. This line of research established that external retrieval could improve knowledge-intensive tasks without requiring every fact to reside in model weights. For organizational deployment, however, retrieval augmentation must be adapted to governance requirements, including approved data sources, document freshness, and evidence attribution. A domain-focused design should therefore regard corpus management and access control as part of the reasoning pipeline rather than as auxiliary engineering.

Open-domain question answering benchmarks provide useful methodological lessons. Natural Questions [6] contains authentic search questions associated with annotated Wikipedia pages, including examples without a supported answer. The benchmark highlights the importance of realistic question distributions, annotation variability, and explicit unanswerability. Domain-specific evaluations should similarly include questions requiring exact lookup, multi-document synthesis, version awareness, and justified abstention. Merely testing a handful of straightforward fact-retrieval questions may inflate performance estimates and conceal failure modes that matter in operational environments.

Hybrid retrieval has become a practical response to the different strengths of lexical and semantic ranking. Reciprocal rank fusion [7] combines ranked lists without requiring directly comparable raw scores, making it useful when BM25 and dense similarity have different numerical scales. Fusion also helps protect exact-match retrieval performance while allowing semantic candidates into the pool. Nevertheless, fused retrieval may still contain redundant, stale, or tangential chunks. Candidate diversity, deduplication, source authority, and temporal metadata are important additional signals when constructing evidence for an LLM.

Cross-encoder reranking offers a more expensive but potentially more discriminative second-stage relevance model. Unlike dual encoders that score independently encoded vectors, a cross-encoder examines the query and candidate passage jointly, permitting fine-grained token interactions. Work on passage reranking with BERT [8] illustrates the effectiveness of such interaction-based ranking. In a domain-specific RAG pipeline, reranking should operate over a bounded candidate set to control inference cost. It should also be tested for robustness to numerical questions and subtle negation, where fluent but incomplete relevance judgments can be misleading.

Document segmentation is another consequential design decision. Fixed-length chunking simplifies indexing but may separate a heading from its governing paragraph, split a definition from its exception, or detach a table from its explanatory notes. Structure-aware chunking attempts to preserve sections, paragraphs, and document hierarchy. Overlap can improve continuity but increases index size and duplicate evidence. A research-quality system should report chunk size, overlap, parsing rules, and the handling of tables and scanned pages because these preprocessing choices influence retrieval metrics as much as the choice of retriever.

Faithfulness and factuality have emerged as distinct quality dimensions for generated answers. An answer can be fluent and responsive while introducing claims unsupported by the supplied documents. RAGAs [4] proposes reference-free metrics to examine retrieval and generation quality, including faithfulness and answer relevance. ARES [5] evaluates context relevance, answer faithfulness, and answer relevance using automated judges calibrated with limited human annotations. These approaches make large-scale testing more feasible, but automated judgments can themselves contain errors. Critical

applications therefore benefit from human auditing, transparent scoring rubrics, and agreement analysis.

Citation attribution is closely related to but not identical with answer correctness. A statement may be true in the broader world yet unsupported by the cited document; conversely, a citation may contain relevant information without establishing the specific claim being made. Reliable systems need claim-level attribution, source identifiers, and ideally passage spans. Verification should assess whether the cited evidence entails the claim, whether a more recent document supersedes it, and whether the answer omits a qualifying condition. For multi-source synthesis, attribution may require more than one citation for a single conclusion.

Recent work on self-reflection and adaptive retrieval has explored decisions about when to retrieve and how to critique generated outputs. Self-RAG [9] proposes reflection tokens to guide retrieval and critique, while corrective RAG [10] emphasizes assessment of retrieved evidence and corrective actions when retrieval quality is inadequate. These directions support a general lesson: retrieval should not be treated as a fixed one-shot operation. However, complex adaptive pipelines may introduce latency, additional model calls, and harder-to-audit behavior. The present framework adopts bounded verification and selective abstention to balance reliability and operational simplicity.

The literature reveals several unresolved integration challenges. A retriever optimized for top-k relevance may not produce the best context for generation; a generator optimized for answer similarity may not be faithful; and a citation system may provide attractive references without proving individual claims. Furthermore, benchmark improvements do not automatically transfer to proprietary documents with noisy formatting and frequent revisions. The proposed architecture addresses these gaps at the system-design level through hybrid ranking, structured metadata, context selection, claim verification, and an evaluation protocol that separates the performance of each stage. Empirical validation remains necessary before claims of superiority can be made.

3. Proposed Methodology

3.1 System architecture and information flow

The system is divided into an offline ingestion/indexing path and an online question-answering path. Approved documents are parsed, normalized, segmented, enriched with metadata, embedded, and indexed. At query time, a permissions-aware search retrieves sparse and dense candidates, fuses rankings, reranks passages, and constructs a bounded evidence context. The generator answers only from this context, and a verification layer checks citation alignment, contradiction, and support before release.

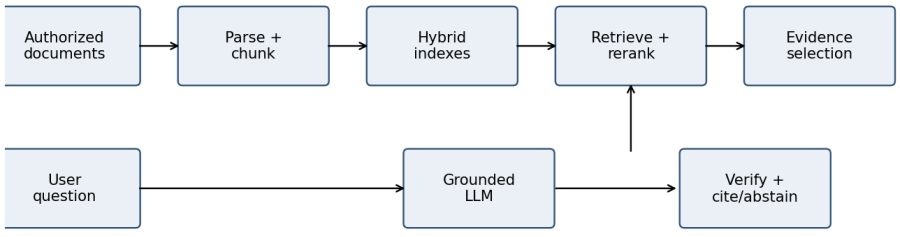


Figure 1: Proposed domain-specific retrieval-augmented question-answering architecture (conceptual design).

3.2 Document ingestion and metadata-aware chunking

Each source document d is assigned an immutable identifier, revision timestamp, owner, access-control label, and source URI. Parsing preserves section titles and table associations where possible. A segmentation operator S divides the document into chunks c_i subject to a token budget B and structural boundaries. Each chunk retains offsets into the original document, permitting later citation verification. The ingestion pipeline rejects malformed records, records extraction errors, and reindexes changed versions.

$$C(d) = S(d; B, o, h), \quad M(c_i) = \{\text{doc_id, section, version, ACL, offsets}\} \tag{1}$$

Here B denotes maximum chunk length, o the overlap allowance, and h structural heading information. Metadata M is stored alongside each chunk; the access-control list (ACL) is applied before candidates can be shown to the generator.

3.3 Hybrid retrieval and reciprocal rank fusion

For query q, a sparse retriever ranks chunks using BM25, while a dense retriever uses cosine similarity between encoded query and passage vectors. Authorized candidates are fused with reciprocal rank fusion, which avoids assuming that lexical and vector scores share a common scale.

$$s_dense(q,c) = (e(q) \cdot e(c)) / (||e(q)|| \, ||e(c)||) \tag{2}$$

$$RRF(c|q) = \Sigma_r \; 1 / (k_0 + rank_r(c|q)) \tag{3}$$

The sum extends over sparse and dense ranked lists; k_0 is a smoothing constant (a preregistered example setting is 60). Candidates absent from a list contribute zero for that list. Duplicate passages are collapsed using document identifiers and offsets.

3.4 Reranking and evidence-budget optimization

A cross-encoder computes a query-dependent relevance score for each candidate. A selection stage balances relevance, source authority, and redundancy while respecting the available prompt budget. This is important because sending too many similar chunks can obscure the decisive evidence and increase latency.

$$s_ce(q,c) = f_\theta([\text{CLS}]\ q\ [\text{SEP}]\ c\ [\text{SEP}]) \tag{4}$$

$$C^* = \text{argmax_C} \; \Sigma(c \in C)[\alpha s_ce(q,c) + \beta a(c)] - \gamma \text{Redundancy}(C) \tag{5}$$

$$\text{subject to } \Sigma(c \in C) \text{ tokens}(c) \le B_context \text{ and } ACL(c,q) = \text{allow}. \tag{6}$$

The parameters α, β, and γ are tuned only on validation data; a(c) measures authority or version priority. Equation (5) represents a constrained selection objective, which may be approximated greedily in implementation.

3.5 Evidence-grounded answer generation

The prompt contains the question, selected evidence with stable citation identifiers, and an instruction to avoid unsupported assertions. The language model produces a draft answer and a mapping between atomic factual claims and supporting passage identifiers. Where a question requests synthesis, the model is permitted to combine evidence only when the conclusion is explicitly supported by the cited passages.

$$\hat{y} = \text{argmax_y} \; P_\varphi(y \mid q, C^*, I_grounded) \tag{7}$$

I_grounded denotes the grounding policy. The generation procedure must preserve qualifiers, numeric units, exceptions, and uncertainty. The model is not instructed to infer facts from an inaccessible source.

3.6 Claim verification and selective abstention

Each answer is decomposed into atomic claims a_j . A verifier estimates whether cited evidence supports each claim and flags missing support or contradiction. The system releases a complete answer only when minimum support and retrieval coverage conditions are met; otherwise it returns a qualified partial answer or abstains.

$$F(\hat{y},C^*) = (1/m)\ \Sigma_j\ \ \text{support}(a_j\ ,\ \text{cited}(C^*)) \tag{8}$$

$$\text{Decision} = \text{answer if } F \geq \tau_f \text{ and coverage} \geq \tau_c\text{; else abstain.} \tag{9}$$

Support is an operational scoring function, not a calibrated probability unless independently validated. Thresholds τ_f and τ_c should be selected on a validation set to satisfy a target error rate. Human review is recommended for high-impact uses.

4. Experimental Design and Evaluation Protocol

4.1 Data, sampling, and baselines

A reproducible study should construct a versioned corpus from an approved specialized domain, such as university regulations, technical manuals, or organizational policies. The corpus should be partitioned at the document-family level to reduce leakage from near-duplicate revisions. Subject-matter experts should annotate a stratified test set containing direct lookup, paraphrased lookup, multi-hop synthesis, conflicting-source, temporal-update, and unanswerable questions. Each item should include a reference answer, supporting document spans, and an answerability label. The proposed comparison comprises: LLM without retrieval; sparse BM25 RAG; dense-only RAG; hybrid RAG; and the full hybrid, reranked, verified framework. All generator settings should be matched across comparable conditions.

4.2 Implementation settings and controls

An implementation may use Python, a document parser, BM25 indexing, a sentence-embedding model, an approximate nearest-neighbor vector index, a cross-encoder reranker, and a fixed-version instruction-following LLM. Model identifiers, prompt templates, chunk size, overlap, retrieval depth, context budget, temperature, seed, hardware, and library versions must be disclosed. As a starting configuration rather than a tested optimum, a researcher could evaluate 400-token chunks with 60-token overlap, 20 candidates per retriever, reranking to 8 passages, and temperature 0. The final values should be chosen on validation data and locked before test evaluation.

4.3 Metrics and statistical analysis

Retrieval should be assessed with Recall@k and mean reciprocal rank (MRR). Answer correctness should be evaluated using exact match or token F1 for short answers and blinded expert scoring for longer responses. Faithfulness is the fraction of answer claims supported by cited passages; citation precision measures the fraction of cited claims whose citations genuinely support them. The abstention rate and selective risk quantify behavior when evidence is insufficient. Median and 95th-percentile latency and resource cost should be reported. Bootstrap confidence intervals over questions and paired significance

tests are recommended; multiple comparisons should be adjusted when appropriate.

$$\text{Recall@k} = (1/N)\ \Sigma_i\ \{\text{relevant evidence in top-k for } q_i\ \} \quad (10)$$

$$\text{MRR} = (1/N)\ \Sigma_i\ 1/\text{rank}_i \quad (11)$$

$$\text{Faithfulness} = \text{supported factual claims} / \text{all factual claims} \quad (12)$$

5. Results and Discussion

IMPORTANT: The following numerical results are hypothetical, constructed solely to demonstrate the required tables, charts, calculations, and interpretation. They are not observations from an implemented experiment and must not be cited as evidence that the proposed system outperforms existing approaches. A submission reporting empirical findings must replace them with measured test-set results.

5.1 Illustrative comparison of system variants

Table 1: Illustrative numerical comparison (synthetic values, not empirical measurements). The example values represent a plausible accuracy–latency trade-off: additional retrieval, reranking, and verification increase computational overhead while potentially improving evidence quality. A real study must report confidence intervals and assess whether improvements persist across domains and question categories.

System	*Recall@5 (%)*	*Answer F1 (%)*	*Faithfulness (%)*	*Citation precision (%)*	*Median latency (ms)*
LLM only	—	64.0	68.0	—	420
BM25 RAG	72.0	73.0	79.0	74.0	650
Dense RAG	78.0	77.0	82.0	78.0	710
Hybrid RAG	85.0	82.0	87.0	84.0	830
Proposed full pipeline	90.0	87.0	93.0	91.0	1120

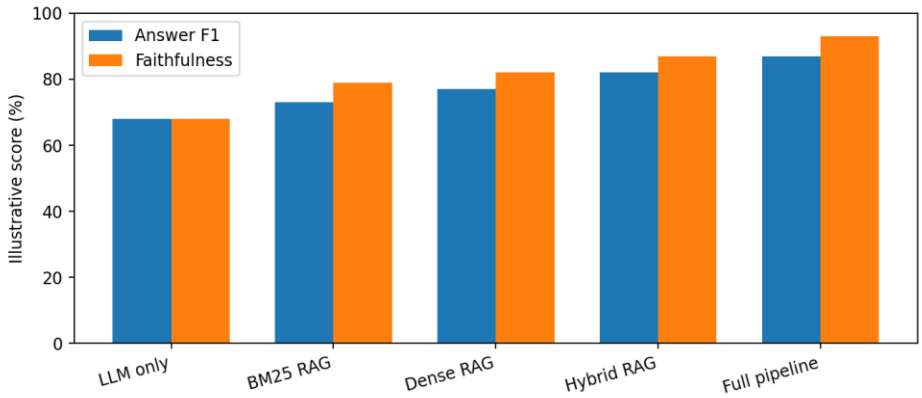


Figure 2: Illustrative (synthetic) answer F1 and faithfulness scores across architectural variants; not measured experimental results.

5.2 Illustrative ablation analysis

Table 2: Hypothetical ablation values for an eventual controlled experiment. The ablation design isolates how lexical retrieval, reranking, verification, and metadata filtering might contribute to overall performance. Importantly, disabling a component should leave all other hyperparameters and test questions unchanged. Real ablation results should be analyzed per question type and should include variance estimates.

Configuration	*Recall@5 (%)*	*Faithfulness (%)*	*Citation precision (%)*
Full pipeline	90.0	93.0	91.0
Without sparse retrieval	82.0	90.0	87.0
Without reranking	86.0	88.0	86.0
Without verification	90.0	82.0	79.0
Without metadata filtering	88.0	85.0	81.0

5.3 Illustrative latency–quality trade-off

Table 3: Hypothetical sensitivity analysis. The illustrative saturation in answer quality at larger context sizes motivates a real optimization experiment: more passages are not necessarily better if they add irrelevant or contradictory material. An appropriate deployment threshold should consider latency budgets, energy use, and the severity of unsupported answers.

Reranked context size	*Answer F1 (%)*	*Faithfulness (%)*	*Median latency (ms)*
Top 2	78.0	89.0	720
Top 4	83.0	91.0	900
Top 8	87.0	93.0	1120
Top 12	87.5	92.0	1470

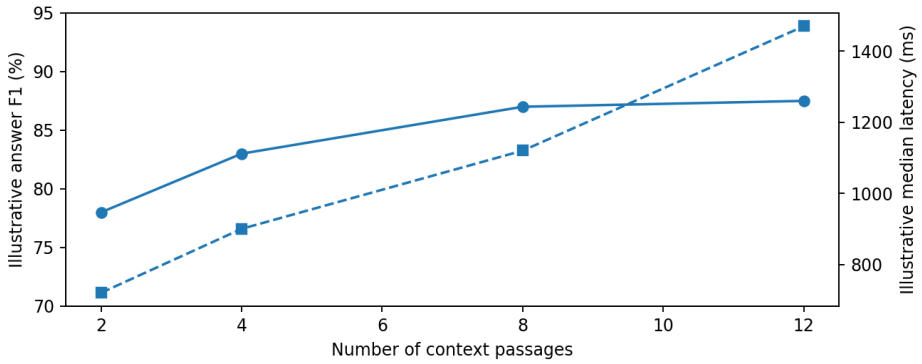


Figure 3: Synthetic sensitivity example showing how quality gains might saturate as latency increases.

5.4 Interpretation, threats to validity, and practical implications

The numerical examples are useful for demonstrating how a future results section should connect retrieval quality to answer quality, but they cannot establish causality or performance. A real experiment must distinguish retrieval misses from generation failures, citation mismatches from factual errors, and

valid abstentions from unnecessary refusals. Dataset construction can introduce selection bias if annotators favor easily searchable questions. Automatically generated questions may overrepresent surface-level facts. Documents may contain sensitive material or embedded malicious instructions; the retrieval layer must enforce authorization, and the generation layer must treat retrieved text as untrusted data rather than executable instructions. Model-based evaluators may favor verbose answers, so blinded expert adjudication is necessary for a representative subset. Temporal drift should be tested by replacing documents with revised versions and measuring whether answers follow the latest authorized source. Domain transfer should be assessed on a second corpus with different vocabulary and document structure.

6. Conclusion

This paper presents a structured retrieval-augmented generative AI framework for domain-specific knowledge discovery and reliable question answering. The proposed architecture integrates metadata-aware ingestion, hybrid lexical–semantic retrieval, cross-encoder reranking, bounded evidence selection, grounded generation, citation verification, and selective abstention. Its mathematical formulation and evaluation protocol make the design implementable and empirically testable. The framework targets three persistent limitations of domain-specific language systems: inadequate retrieval of authoritative evidence, unsupported generation, and weak source traceability. Numerical tables and figures were provided as explicitly hypothetical reporting examples, not experimental findings; accordingly, no empirical performance improvement is claimed. Future work should implement the framework on a permissioned, versioned corpus, compare it against matched baselines, report uncertainty intervals, test robustness to document revisions and prompt injection, and quantify the trade-offs among correctness, faithfulness, abstention, latency, and cost. Reliable deployment will depend on the combined performance of retrieval, generation, verification, governance, and ongoing human evaluation.

References

Mahadasu, S. S. (2024). Closed-loop sustainability orchestration for energy-aware RAN, edge, and data-center workloads in 5G-Advanced networks. International Journal of Engineering Research and Science & Technology, 20(2), 1392–1398. https://doi.org/10.62643/ijerst.2024.v20.n2.4407

Anchala, S. (2019). The evolution of DevSecOps in software supply chain security: Addressing dependencies, open-source vulnerabilities, and the prevention of software supply chain attacks. International Journal of Multidisciplinary and Scientific Emerging Research, 7(1), 415–422. https://doi.org/10.15662/IJMSERH.2019.0701047

Aggarwal, A. (2025). Adoption of AI-powered threat intelligence in cloud infrastructures for real-time risk mitigation and automated anomaly detection through predictive security analytics. Journal of Digital Security and Forensics, 2(1), 139–150. https://doi.org/10.29121/digisecforensics.v2.i1.2025.87

Jain, Arpit, Nageswara Rao Moparthi, A. Swathi, Yogesh Kumar Sharma, Nitin Mittal, Ahmed Alhussen, Zamil S. Alzamil, and MohdAnul Haq. "Deep Learning-Based Mask Identification System Using ResNet Transfer Learning Architecture." Computer Systems Science & Engineering 48, no. 2 (2024): 341.

Bhaskara Raju Rallabandi. (2023). AGENTIC-AI xAPPS: O-RAN MULTI-AGENT CONTROL FOR ENTERPRISE POLICY, SCHEDULING & SPECTRUM. International Journal of Data Science and IoT Management System, 2(4), 54–58. https://doi.org/10.64751/ijdim.2023.v2.n4.pp54-58

Ahuja, R. (2019). A comparative study of traditional DevOps and DevSecOps: Examining the cultural, technical, and organizational shifts required to embed security into software engineering practices. International Journal of Innovative Research in Science, Engineering and Technology, 8(8), 8561–8569. https://doi.org/10.15680/IJIRSET.2019.0808096

Jain, Arpit, Adesh Kumar, and Sanjeev Sharma. "Comparative design and analysis of mesh, torus and ring NoC." Procedia Computer Science 48 (2015): 330–337.

Devulapalli, N. (2020). The evolution of database security from centralized architectures to distributed cloud-native systems: A comparative and historical perspective. Turkish Online Journal of Qualitative Inquiry, 11(4), 2954–2963. https://doi.org/10.69980/j1064z97

Mahadasu, S. S. (2022). Cross-domain service assurance for private 5G networks using streaming telemetry and dependency-aware event correlation. International Journal of Data Science and IoT Management System, 1(3), 51–56. https://doi.org/10.64751/ijdim.2022.v1.n3.1290

Harshitha, G., Kumar, S., Rani, S., & Jain, A. (2021, November). Cotton disease detection based on deep learning techniques. In 4th Smart Cities Symposium (SCS 2021) (Vol. 2021, pp. 496–501). IET.

Devulapalli, N. (2023). Database security and big data analytics: Protecting large-scale data warehouses, ensuring query privacy, and mitigating emerging threats in real-time analytics systems. International Journal on Recent and Innovation Trends in Computing and Communication, 11(7), 622–629. https://www.ijritcc.org/index.php/ijritcc/article/view/11939

Bhaskara Raju Rallabandi. (2022). SYNC-ADAPTIVE RESOURCE SCHEDULING IN PRIVATE 5G FOR TIME-CRITICAL APPLICATIONS. International Journal of Data Science and IoT Management System, 1(2), 40–45. https://doi.org/10.64751/ijdim.2022.v1.n2.pp40-45

Ahuja, R. (2020). Adoption of secure API development practices in microservice architectures for preventing unauthorized access and data leakage through OAuth2 and input throttling. Journal of Critical Reviews, 7(9), 4063–4069. https://doi.org/10.69980/jcr.v7:i9.13404

Khan, S. (2022). The use of artificial intelligence in ESG (environmental, social, and governance) investing: A study on AI models for sustainability analytics in asset and wealth management. International Journal of Multidisciplinary and Scientific Emerging Research, 10(1), 431–439. https://doi.org/10.15662/IJMSERH.2022.1001046

Aggarwal, A. (2021). Utilization of continuous vulnerability scanning in DevSecOps pipelines for preemptive threat detection and secure code deployment through automated security checks. NeuroQuantology, 19(1), 427–442. https://doi.org/10.48047/nq.2021.19.1.NQ21055

Jain, Jaishree, Arpit Jain, Saurabh Kumar Srivastava, Chaman Verma, Maria Simona Raboaca, and Zoltán Illés. "Improved security of e-healthcare images using hybridized robust zero-watermarking and hyper-chaotic system along with RSA." Mathematics 10, no. 7 (2022): 1071.

Anchala, S. (2020). Cloud security challenges in supply chain systems: Addressing vendor risks, data integrity, and third-party service provider vulnerabilities. Journal of Critical Reviews, 7(2). https://doi.org/10.69980/jcr.v7:i2.13401

Ahuja, R. (2023). The role of artificial intelligence and machine learning in DevSecOps: Leveraging predictive analytics, automated threat detection, and anomaly identification for secure software delivery pipelines. International Journal on Recent and Innovation Trends in Computing and Communication, 11(2), 381–389. https://www.ijritcc.org/index.php/ijritcc/article/view/11946

Devulapalli, N. (2018). Database security for financial systems: Protecting customer data, ensuring PCI-DSS compliance, and safeguarding against fraud and insider attacks. International Journal of Advanced Research in Education and Technology, 5(5), 838–844. https://doi.org/10.15680/IJARETY.2018.0505018

Dwivedi, D. (2018). The role of security champions in DevSecOps teams: Empowering developers with security skills and bridging the gap between security specialists and

engineering teams. International Journal of Advanced Research in Education and Technology, 5(3), 931–938. https://doi.org/10.15680/IJARETY.2018.0503016

Mahadasu, S. S. (2024). AI-native wireless digital twins for precision agriculture: Edge intelligence and predictive farm operations in 6G ecosystems. International Journal of Data Science and IoT Management System, 3(1), 65–72. https://doi.org/10.64751/ijdim.2024.v3.n1.1287

Kumar, Ashok, and Arpit Jain. "Image smog restoration using oblique gradient profile prior and energy minimization." Frontiers of Computer Science 15, no. 6 (2021): 156706.

Anchala, S. (2020). Adoption of secure access service edge (SASE) in distributed enterprises for ensuring cloud application protection and network optimization through unified security frameworks. Turkish Online Journal of Qualitative Inquiry, 11(4), 2964–2973. https://doi.org/10.69980/60m26411

Rao, S. Madhusudhana, and Arpit Jain. "Advances in Malware Analysis and Detection in Cloud Computing Environments: A Review." International Journal of Safety & Security Engineering 14, no. 1 (2024).

Khan, S. (2018). The role of zero-trust models in database security: Eliminating implicit trust and enforcing continuous verification in enterprise data access systems. International Journal of Research in Electronics and Computer Engineering, 6(1), 1622–1628.

Gupta, Namit, Kunwar Singh Vaisla, Arpit Jain, Adesh Kumar, and Rajeev Kumar. "Performance Analysis of AODV Routing for Wireless Sensor Network in FPGA Hardware." Computer Systems Science & Engineering 40, no. 3 (2022).

Khan, S. (2019). Sales force security compliance: An in-depth study of GDPR, HIPAA, and PCI-DSS enforcement in cloud-based CRM systems and their implications for global enterprises. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 8(9), 2211–2219. https://doi.org/10.15662/IJAREEIE.2019.0809010

Kumar, S., Sharma, K., Kumar, P. A., Jain, A., Bhagat, S. K., & Singh, P. (2024, September). An improved particle swarm approach for energy-aware location-aided routing in mobile ad-hoc network. In 2024 7th International Conference on Contemporary Computing and Informatics (IC3I) (Vol. 7, pp. 1119–1124). IEEE.

Anchala, S. (2024). Deployment of zero trust security models in hybrid cloud infrastructures for minimizing lateral movement and strengthening access control through continuous verification mechanisms. International Journal of Innovative Research in Computer and Communication Engineering, 12(7), 9993–10002. https://doi.org/10.15680/IJIRCCE.2024.1207114

Bhaskara Raju Rallabandi. (2024). DESIGNING NTN–TERRESTRIAL 6G EXPERIMENTS FOR AGRICULTURE AND SMART GRIDS. American Journal of Management and IOT Medical Computing, 3(3), 27–32. https://doi.org/10.64751/ajmimc.2024.v3.n3.pp27-32

Ahuja, R. (2020). Database security in smart cities: Protecting interconnected infrastructure data, ensuring citizen privacy, and managing large-scale IoT data streams. International Journal of Multidisciplinary and Scientific Emerging Research, 8(1), 410–418. https://doi.org/10.15662/IJMSERH.2020.0801047

Kumar, S., Jain, A., & Moparthi, N. R. (2022). Enhanced method of object tracing using extended Kalman filter via binary search algorithm.

Mahajan, R. A., Dey, R., Khan, M., Mohd Su'ud, M., Alam, M. M., & Jadhav, P. (2025). Enhancing personalization in IoT-based health monitoring via generative AI and transfer learning. Egyptian Informatics Journal, 32, Article 100788. https://doi.org/10.1016/j.eij.2025.100788

Bhaskara Raju Rallabandi. (2024). 6G Cross-Layer Intent Engine: Automated Translation from Business Requirements to Network Policies and Real-Time Scheduling Controls. International Journal of Data Science and IoT Management System, 3(1), 44–48. https://doi.org/10.64751/ijdim.2024.v3.n1.pp44-48

Aggarwal, A. (2024). Adoption of AI-powered fraud detection in online transactions for enhancing financial security and trust through pattern-based anomaly detection

techniques. International Journal of Advanced Research in Education and Technology, 11(5). https://doi.org/10.15680/IJARETY.2024.1105043

Jain, A. (2024, December). Energy-Efficient Signal Processing Architectures for IoT Networks Design. In 2024 International Conference on Emerging Technologies and Innovation for Sustainability (EmergIN) (pp. 362–367). IEEE.

Mahajan, R. A., Dey, R., Khan, M., Kanabar, K., & Gupta, S. K. (2025). Real-time vehicular-to-vehicular (V2V) communication and applications for pedestrian safety in AI-optimized VANETs for autonomous vehicles. In H. Maryam, M. M. Malik, I. U. Khan, & S. K. Gupta (Eds.), AI-driven transportation systems: Real-time applications and related technologies (pp. 151–167). Springer. https://doi.org/10.1007/978-3-031-98349-8_8

Mahadasu, S. S., & Rallabandi, B. R. (2023). Federated learning for privacy-preserving RAN optimization across multi-site enterprise 5G networks. International Journal of Engineering Science and Advanced Technology, 23(4), 53–57. https://doi.org/10.64771/ijesat.2023.v23.i04.pp3847

Aggarwal, A. (2025). Artificial intelligence in criminal justice systems: An examination of predictive policing, automated surveillance, and ethical concerns in legal decision-making. International Journal of Multidisciplinary Research in Science, Engineering and Technology, 8(1). https://doi.org/10.15680/IJMRSET.2025.0801070

Bhaskara Raju Rallabandi. (2023). CLOSED-LOOP DIGITAL-TWIN MANAGEMENT FOR HYBRID 5G-NTN NETWORKS. American Journal of Management and IOT Medical Computing, 2(3), 30–35. https://doi.org/10.64751/ajmimc.2023.v2.n3.pp30-35