POLICYRAG: PROMPT-GUIDED SYMBOLIC GRAPH MEMORY FOR INTERPRETABLE MULTI-HOP RETRIEVAL

Anonymous authors
Paper under double-blind review

ABSTRACT

Retrieval augmented generation is a powerful way to ground large language models in external knowledge, yet most pipelines still treat the prompt as instructions for text production rather than as a control surface for retrieval. We introduce PolicyRAG, a framework that recasts retrieval as an explicit, auditable policy operating over a symbolic graph memory. Text is organized into lightweight typed links between entities and passages, enabling transparent search and controllable evidence selection. Beyond accuracy, the policy is compact and human editable, supporting governance, domain adaptation, and safety review without retraining. At query time, the policy seeds candidate entities, invokes brief LLM calls only for disambiguation and local gating, and performs symbolic traversal with Personalized PageRank (PPR). The resulting scores are projected to passages and finalized with a small, transparent re-ranker, producing a per-query trace of seeds, paths, and scores for explainable evidence selection. Compared with long-context expansion, the policy keeps test-time compute modest while preserving answer quality. On multi-hop question answering benchmarks, PolicyRAG achieves state-of-the-art results on HotpotQA (F1 80.7), 2WikiMultiHopQA (F1 78.9), and MuSiQue (F1 55.9) while remaining fully auditable and training free. We also assess domain adaptability on domain-specific datasets. By coupling symbolic structure with prompt level control, PolicyRAG provides a practical route from question to verifiable evidence and advances accuracy, efficiency, and trust in retrieval augmented generation.

1 INTRODUCTION

Large language models (LLMs) have improved language comprehension and generation in natural language processing. Advanced neural networks exhibit capabilities in text completion, summarization, reasoning, and code development. Despite high test scores, their reliability is questioned, particularly in tasks that necessitate moving beyond pattern recognition to synthesize dynamic knowledge, integrate information from various sources, and present logically coherent and verifiable evidence. The inability to access real-time or specific domain knowledge without training highlights fundamental constraints. Typically, large language models (LLMs) function as static repositories, reflecting learned data patterns without the capacity for updates or new data integration during inference. This results in knowledge deterioration, factual errors, and verification challenges, which affect applications requiring high accuracy. Retrieval-augmented generation (RAG) frameworks address these issues by integrating LLMs with external data sources, while dense passage retrieval techniques based on semantic similarity improve RAG systems by enabling live data integration, minimizing errors, and providing precise source referencing Karpukhin et al. (2020); Lewis et al. (2020b). RAG exhibits competence in resolving single-hop issues within a single document; however, it encounters difficulties with multi-hop reasoning that entails concealed entities or causal relationships. Multi-hop thinking requires gathering data from diverse sources, recognizing hidden connections, and synthesizing information in a coherent and accurate manner. Traditional RAG shows limitations in multi-hop QA benchmarks, as indicated by performance disparities between humans and machines on multi-document tasks, particularly in the HotpotQA datasets Yang et al.

(2018). Databases like 2WikiMultiHopQA face similar challenges Ho et al. (2020), while MuSiQue emphasizes retrieval limitations in complex scenarios based on embeddings Trivedi et al. (2022).

Proposed solutions to these challenges include the use of graph-driven methods for thorough multi-hop inference. Graph-based methods convert unstructured data into structured formats, facilitating a comprehensive analysis of relationships and dependencies Asai et al. (2020); Sun et al. (2018). These methodologies utilize graph models, which outperform dense retrieval methods in multi-hop tasks by clearly delineating relational structures. This functionality clarifies the relationships between entities, the chronological progression of events, and the causal connections that may be obscured when relying solely on the similarities present in embeddings. The incorporation of explainable artificial intelligence enhances the transparency of analyses by facilitating interpretation, verification, and problem resolution within specific domains of expertise. Structured exploration of graphs provides a more reliable approach for enhancing compositional reasoning, as opposed to the potentially unreliable reliance on embedding similarities. It is anticipated that recent developments in reasoning systems that are impacted by neurobiology will increase processing efficiency, which will result in a significant enhancement of multi-hop reasoning skills being achieved. In order to create associative memory networks, HippoRAG2 makes use of hippocampal indexing theory. This results in increased accuracy and reduced computing requirements in comparison to typical iterative retrieval techniques Gutiérrez et al. (2025). However, current techniques often use prompts as generating aids rather than as tools for managing retrieval. This is despite the fact that there is evidence to show that symbolic structures are necessary for sophisticated cognitive processes.

The currently available graph-based approaches have drawbacks since they primarily see interactions between language models as creative activities. As a result, they lose out on possibilities to enhance retrieval procedures. Conventionally, during post-retrieval stages, systems tend to prioritize generation-oriented prompts, mainly neglecting the dynamic retrieval supervision and evidence filtering capabilities that are inherent to language models. An obstacle of this kind makes it difficult to modify retrieval algorithms so that they are customized to the particular requirements of individual queries, domain-specific features, and evidence quality. A policy-oriented prompting technique is proposed in this study as a means of overcoming these restrictions and redefining the integration of language models with structured knowledge systems. Through strategic observation of prompt policies and their effective implementation, the framework conceptualizes each query as a guiding policy that dictates the selection of initial entities, sets standards for the retention of localized facts, formulates strategies for navigating typed graphs, and fine-tunes evidence for generative processes. This shift progresses from elementary prompt-based generation to sophisticated retrieval regulation through employing cohesive symbolic memory architectures. The advocated method underscores an enhanced symbolic memory framework that cohesively integrates entities, snippets of text, and categorized associations under a refined control system. This integration maintains the clarity of graph-based reasoning, boosts retrieval efficacy, and optimizes evidence usage. Consequently, the approach minimizes superfluous content while emphasizing logical reasoning routes, culminating in concise and robust evidence compilations that facilitate the generation of trustworthy and substantiated answers.

This work makes specific contributions:

1. We introduce PolicyRAG, a training-free graph-based retrieval system that implements policy-driven prompting through five key mechanisms: entity extraction with relevance ranking, keep-drop prompting for essential information filtering, Personalized PageRank evidence compilation, multi-factor passage scoring, and structured re-ranking with title concordance and seed integration.

2. We demonstrate that effective retrieval control can be achieved through symbolic memory controllers that operate without extensive training, employ LLMs selectively, and maintain complete decision audit trails for transparency and verification.

3. We report state-of-the-art performance on multi-hop question answering benchmarks, with F1 scores on HotpotQA, 2WikiMultiHopQA, and MuSiQue datasets, while maintaining superior early-rank recall and interpretable reasoning pathways. Yang et al. (2018); Ho et al. (2020); Trivedi et al. (2022).

2 Related Work

2.1 Multi-hop QA with LLMs and Graphs

The task of answering multi-hop questions highlights a significant challenge in modern language comprehension systems. This complexity arises because crucial information is distributed across several documents, and these pieces are linked solely by hidden entity associations. Although large language models are proficient in handling single-step reasoning tasks, their effectiveness diminishes considerably when solutions necessitate the integration of information from multiple sources, supported by traceable chains of evidence Yang et al. (2018); Ho et al. (2020); Trivedi et al. (2022). Conventional retrieval-augmented generation frameworks encounter difficulties with multi-hop queries, as relying on embedding similarity is inadequate for capturing the complex associations that are essential for the reasoning needed to solve questions involving multiple steps Karpukhin et al. (2020); Lewis et al. (2020b).

Graph-Enhanced and Structure-Augmented RAG. Techniques strengthened by graph structures provide comprehensive knowledge graphs, which permit the discovery of relationships among diverse entities and passages, hence boosting retrieval efficacy and the trustworthiness of the acquired answers. The QA-GNN framework laid down essential methodologies for integrating reasoning across textual environments with structured knowledge graphs, accomplished through intricate dual-component system architectures Yasunaga et al. (2021). The advancements of GreaseLM in achieving state-of-the-art outcomes were driven by its innovative integration of text and graph encodings Zhang et al. (2022a). In contrast, Think-on-Graph posits large language models as independent entities navigating and interpreting knowledge graph connections without the requirement for retraining Sun et al. (2024). Current approaches that augment structures encompass RAPTOR, which employs recursive clustering to construct and distill text segments into hierarchical tree formations, achieving notable performance gains, reaching up to 20% on the QuALITY benchmark Sarthi et al. (2024). Furthermore, GraphRAG composes elaborate knowledge graphs specifically tailored for comprehensive reasoning across graph schematics Edge et al. (2024), whereas LightRAG focuses on enhancing the efficiency of graph-based information retrieval processes in a computationally efficient manner Guo et al. (2024). Collectively, these innovative strategies address the drawbacks inherent in conventional retrieval systems, which traditionally handle only short, contiguous pieces of text.

Neurobiologically-Inspired Memory Systems. Architectures drawing inspiration from neurobiology replicate principles observed in the hippocampus concerning memory formation and associative recall mechanisms. Through the application of hippocampal indexing theory, the model known as HippoRAG achieved notable advancements, such as a 20% enhancement in performance while concurrently reducing computational costs by a factor of 10 to 30 Gutiérrez et al. (2024). Building on this foundation, HippoRAG2 realized further refinement by attaining an additional 7% improvement in performance metrics, all while maintaining a robust comprehension of factual information. This empirical evidence underscores that employing symbolic frameworks can enhance accuracy significantly, alongside a decrease in computational resource requirements Gutiérrez et al. (2025). Additionally, the approaches termed Generate-on-Graph and Graph Chain-of-Thought underscored the effectiveness of integrating neural and symbolic reasoning paradigms, thereby reinforcing the potential advantages of hybrid systems Xu et al. (2024); Jin et al. (2024).

2.2 Policy-Driven Retrieval and Auditable Control

Recent research has identified the need for policy-driven prompting methodologies that envision each question as a structured policy dictating seed selection, local fact retention, typed graph navigation, and evidence presentation Wei et al. (2022); Yao et al. (2022; 2023). This paradigmatic shift moves beyond traditional prompt-based generation toward systematic retrieval control through symbolic memory frameworks comprising entities, passages, and typed connections managed by minimalistic controllers Weston et al. (2014); Andreas et al. (2016). The policy-driven approach implements enduring symbolic memory architectures where seeds derive from question analysis, proximate facts undergo filtration through keep-drop prompting Dhuliawala et al. (2023), Personalized PageRank traversals compile evidence across relational linkages Page et al. (1999); Jeh & Widom (2003), and re-ranking processes emphasize title alias concordance and multi-seed coverage Nogueira & Cho (2019). These mechanisms operate without extensive training, employ LLMs

judiciously, and maintain comprehensive audit trails for verification Wei et al. (2022); Zhang et al. (2022b).

A complementary strand pursues test-time control instead of scaling decoders or context windows. Self-critique methods close the retrieval–generation loop by letting a model accept or reject its own evidence, but typically rely on a learned critic and extra training Asai et al. (2023). Surveys of graph-based RAG make a strong case for symbolic structure as a route to transparency and easy editing, yet many pipelines still couple retrieval with heavy summarization stages or learned rerankers Peng et al. (2024). Our stance is lighter a policy controller over a typed graph surfaces, per query, the restart mass, traversal mix, and scoring rules, and can steer exploration in the spirit of topic-sensitive PageRank while remaining straightforward to audit Haveliwala (2002). This situates our work within a broader turn toward editing-first, auditable retrieval, where gains come from explicit control of evidence paths rather than ever-longer contexts.

3 PolicyRAG Framework

We treat the prompt as a retrieval policy that operates over a symbolic graph memory. The framework has two regimes. In Symbolic Memory Construction, we compile an entity–centric memory with typed connectivity and stable provenance. In Query–Time Policy Execution, each question instantiates a lightweight controller that seeds entities, gates nearby facts, traverses the graph, scores passages, and synthesizes an answer, as shown in Figure 1. The design is model agnostic and focuses on concrete pipeline steps and tunable knobs that deliver stability, efficiency, and traceable evidence without committing to specific extractors, encoders, or generators.

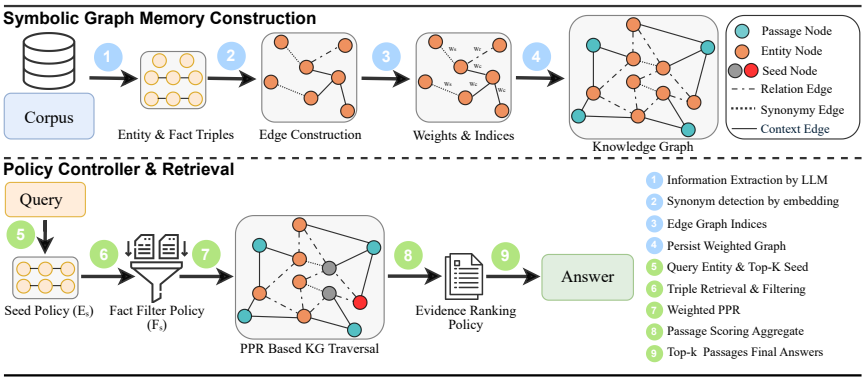

Figure 1: **Policy-RAG methodology.** Steps include offline indexing and memory construction through entity and fact triple extraction, weighted graph persistence, knowledge graph construction, seed selection and fact filtering, traversal via Personalized PageRank (PPR), passage scoring and reranking, and final answer synthesis.

Notation. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a heterogeneous graph with entity nodes $\mathcal{V}_E$, passage nodes $\mathcal{V}_P$, and three edge families: context E_c, relation E_r, and synonymy E_s. Let A_c, A_r, A_s denote the corresponding row–stochastic transition operators on entities, and let $B \in \{0,1\}^{|\mathcal{V}_E| \times |\mathcal{V}_P|}$ be the entity–to–passage incidence. Denote the probability simplex in by Δ^n.

3.1 Symbolic Graph Memory Construction

Source articles are normalized and segmented into titled passages with stable identifiers, byte offsets, and provenance pointers, producing a uniform corpus that supports auditable lookups. Salient entities are discovered and canonicalized; surface forms, redirects, and aliases are consolidated so that mentions resolve to stable entity identifiers. Optional factual assertions are extracted as triples (s, p, o) and anchored to the passages from which they were derived.

Graph schema and typed connectivity. We materialize a heterogeneous memory $\mathcal{G}$ with two node types (entities, passages) and three edge families: E_c for context links from entities to passages

(mentions), E_r for typed relations between entities, and E_s for synonymy or alias edges between entities. Separating these families preserves route semantics, enables type–wise weighting during traversal, and keeps the memory lightweight. Triples that do not resolve to canonical entities are discarded; duplicate or low–confidence edges are pruned.

Indices and persistence. For scalable access, we persist sparse adjacency lists per edge type, row–normalized transition operators, and lookup caches. Lightweight priors used later for normalization are precomputed (passage length, entity specificity via inverse passage frequency). Edge–type mixture weights (η_c, η_r, η_s) and segmentation parameters are fixed per corpus and reused across experiments. The graph and indices are stored in a persistent backend and loaded read–only at query time.

Composite transitions. With row normalized A_τ, we define block transitions

$$T_E \;=\; \eta_r\, A_r \;+\; \eta_s\, A_s, \qquad T_{E\to P} \;=\; \text{row-norm}(B), \qquad T_{P\to E} \;=\; B^\top, \tag{1}$$

so T_E moves within the entity layer via relations and synonymy, $T_{E\to P}$ aggregates entity mass to passages through context, and $T_{P\to E}$ can optionally re–inject passage mass into entities for iterative schemes.

3.2 Query–Time Policy Controller

Given a question q, a prompt–level policy governs seeding, local fact gating, typed Personalized PageRank (PPR), normalization, and re–ranking. The controller is lightweight and training–free.

Seed policy. We form a small seed set $\mathcal{E}_s$ and a restart distribution $\mathbf{s} \in \Delta^{|\mathcal{V}_E|}$:

$$\mathbf{s}_e \;=\; \frac{\text{align}(e,q)\cdot \text{spec}(e)\cdot \text{alias\_ok}(e)}{\sum_{e'} \text{align}(e',q)\,\text{spec}(e')\,\text{alias\_ok}(e')}\,, \tag{2}$$

Here, $\text{align}(e,q)$ can be lexical or embedding–based similarity, $\text{spec}(e) \in [0,1]$ down–weights generic heads, and $\text{alias\_ok}(e)$ suppresses over–broad aliases.

Local fact gating. We collect candidate triples near $\mathcal{E}_s$ and retain a compact set $\mathcal{F}_s$ using a short keep–drop prompt:

$$\mathcal{F}_s \;=\; \{\, f \in \text{Cand}(\mathcal{E}_s) \;:\; \text{keep}(f,q) = 1 \,\}. \tag{3}$$

Exploration then prefers neighborhoods consistent with $\mathcal{F}_s$, which reduces distractors without expanding compute.

Typed traversal (PPR). We walk over the entity subgraph with a type mixture

$$T_E \;=\; \eta_r\, A_r \;+\; \eta_s\, A_s, \qquad \eta_r, \eta_s \ge 0,\; \eta_r + \eta_s = 1, \tag{4}$$

and run PPR with damping $\alpha \in (0,1)$:

$$\mathbf{v}^{(t+1)} \;=\; (1-\alpha)\, T_E^\top \mathbf{v}^{(t)} \;+\; \alpha\, \mathbf{s}, \qquad \mathbf{v}^{(0)} = \mathbf{s}, \quad \|\mathbf{v}^{(t+1)} - \mathbf{v}^{(t)}\|_1 < \varepsilon \text{ or } t = \tau_{\max}. \tag{5}$$

We denote the fixed point by $\mathbf{v}^\star$ concentrates on entities that are aligned with q and supported by typed topology, relational hops vs. alias consolidation.

Projection and normalization. Entity scores are projected to passages via the context incidence, $\mathbf{u} = B^\top \mathbf{v}^\star$, then normalized to mitigate length/frequency bias:

$$\tilde{u}_p = \frac{u_p}{\ell(p)^\beta\, \text{IPF}(p)^\gamma}, \qquad \beta, \gamma \in [0,1], \tag{6}$$

where $\ell(p)$ is passage length and $\text{IPF}(p)$ denotes specificity.

Re–ranking and evidence selection. Final scores add small, transparent bonuses for title/alias hits, multi–seed coverage, and short, type–coherent paths:

$$S(p \mid q) \;=\; \tilde{u}_p \;+\; \lambda_{\text{title}}\, \phi_{\text{title}}(p,q) \;+\; \lambda_{\text{cov}}\, \phi_{\text{cov}}(p, \mathcal{E}_s) \;+\; \lambda_{\text{path}}\, \phi_{\text{path}}(p), \tag{7}$$

and $\text{TopK}_p\, S(p \mid q)$ defines the evidence used for answer synthesis. All components are deterministic the complete trace (seeds, kept facts, paths, ranks, scores) is logged per query.

4 EXPERIMENTAL SETUP

4.1 BASELINES

We compare against three complementary families under a single, controlled protocol that fixes the passage inventory, dataset splits, and inference budget. First, to anchor results in established retrieval practice, we include lexical and dense retrievers BM25 for sparse term matching alongside widely used dense encoders such as DPR, Contriever, and GTR, and a standard RAG pipeline that couples dense retrieval with generation Robertson & Walker (1994); Lewis et al. (2020a); Izacard et al. (2021); Ni et al. (2021). Second, to reflect modern retrieval capacity independent of graph structure, we add large embedding models that report strong BEIR performance, including GTE-Qwen2-7B-Instruct, GritLM-7B, and NV-Embed-v2 Thakur et al. (2021); Li et al. (2023); Muennighoff et al. (2024); Lee et al. (2024). Third, to test whether explicit connectivity improves evidence finding and answerability under the same compute envelope, we evaluate structure-augmented RAG methods RAPTOR, which organizes the corpus hierarchically GraphRAG and LightRAG, which exploit graph structure for retrieval and summarization and HippoRAG and HippoRAG 2, which integrate a knowledge graph with diffusion-style scoring Sarthi et al. (2024); Edge et al. (2024); Guo et al. (2024); Jimenez Gutierrez et al. (2024); Gutiérrez et al. (2025). All baselines operate on the identical passage pool with matched token and latency budgets when public checkpoints and recommended hyperparameters are available we use them, otherwise we retain the reported configurations, with full settings provided in Appendix E. To disentangle retrieval quality from context length, we also report a long-context control that concatenates the same retrieved passages to the generator without any graph reasoning.

4.2 DATASETS

We evaluate on three established multi-hop question-answering benchmarks that probe complementary aspects of complex reasoning, each constructed from or aligned to Wikipedia[1]. This choice ensures a common provenance while stressing different retrieval and composition behaviors that matter for graph-enhanced RAG. **HotpotQA** Yang et al. (2018) contains questions that require evidence drawn from multiple articles, prominently including bridge and comparison types. Each question is paired with gold supporting sentences, enabling faithful evaluation of multi-document reasoning and evidence use. **2WikiMultiHopQA** Ho et al. (2020) is built from curated pairs of Wikipedia pages and emphasizes clean two-hop chains across entities and pages its construction reduces lexical shortcuts and encourages genuine cross-article inference.

Table 1: Dataset statistics.

	NQ	PopQA	2Wiki	HotpotQA	MuSiQue
Number of Queries	1,000	1,000	1,000	1,000	1,000
Number of Passages	9,633	8,676	9,811	6,119	11,656

MuSiQue Trivedi et al. (2022) composes questions from single-hop sub-questions but injects strong distractors, stressing keep drop decisions and robustness to off-topic yet topically related content. For completeness, Table 1 also lists corpus sizes for **NQ** Wang et al. (2024) and **PopQA** Mallen et al. (2022), which are commonly used for single-hop factual QA. We include them to standardize corpus statistics, our primary multi-hop evaluation focuses on HotpotQA, 2Wiki, and MuSiQue. An illustrative example is provided in Figure 2

Domain-specific corpora. To assess domain adaptability without retraining, we additionally evaluate domain datasets in Legal, Health, and History details in Appendix C. These corpora provide out-of-distribution terminology, citation styles, and entity linking patterns, enabling a targeted stress test of policy editability, governance, and transfer.

[1] https://www.wikipedia.org/

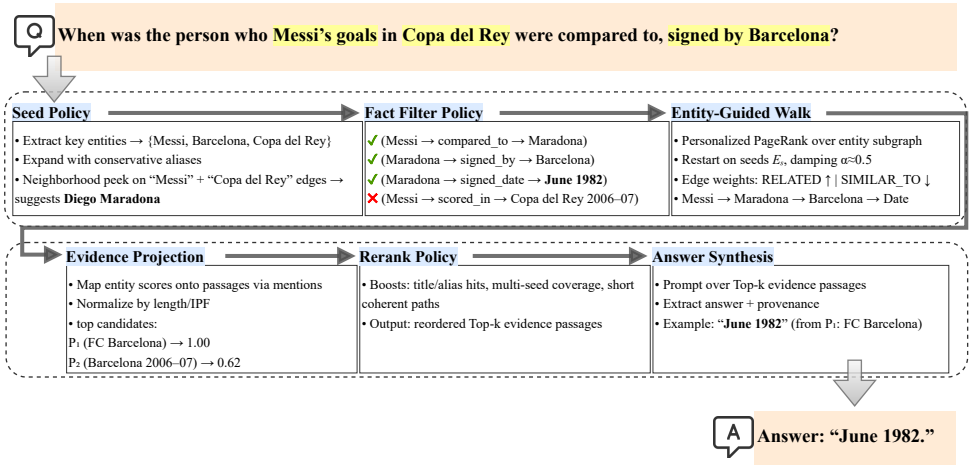

Figure 2: Example of the Policy-Driven Retrieval-Augmented Generation (RAG) framework for multi-hop question answering.

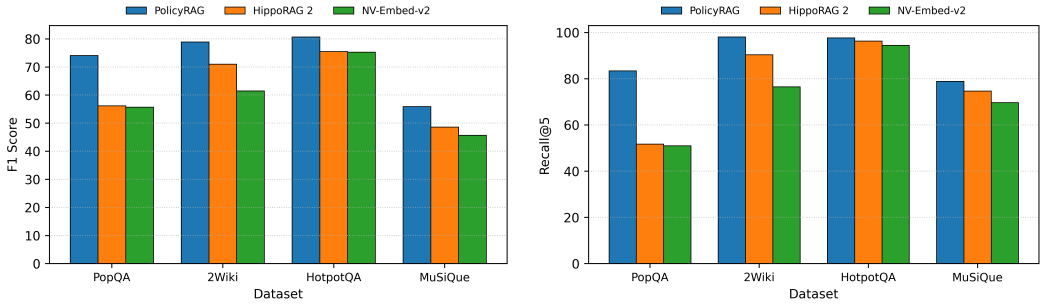

Figure 3: PolicyRAG performance: F1 Score and Recall@5 across PopQA, 2Wiki, HotpotQA, and MuSiQue.

4.3 Evaluation Metrics

Our primary answer metric is token-level F1 on normalized strings, following prior multi-hop QA work Yang et al. (2018); Ho et al. (2020); Trivedi et al. (2022), and we also report exact match (EM) for completeness. Retrieval quality is measured by recall@k with an emphasis on early ranks, and is complemented by MRR and nDCG to capture the quality of ordering. Beyond these task metrics, we track diagnostics that assess whether the controller is selecting evidence that actually supports the answer rather than simply accumulating more text answer-supported@k is the fraction of questions whose gold answer is derivable from the top-k passages without external context HopsRecovered@k measures how often the entity chain implied by the gold rationale is fully discoverable within the top-k and the Distractor Suppression Rate quantifies the proportion of retrieved items that are subsequently rejected by the local fact gate near the seeds. To assess robustness, we also measure the stability of ranked lists under small paraphrases and minor alias substitutions in the query. Taken together, these metrics expose precision coverage trade-offs and the specific contribution of typed traversal and gating, complementing headline F1 and EM and aligning with recent evaluations of structure-augmented RAG.

4.4 Implementation Details

PolicyRAG is training-free and configuration-light across all datasets. We build the symbolic memory once per corpus (Wikipedia-aligned) and load it read-only at inference. At query time, the controller applies a compact seed policy, a short keep–drop prompt for local fact gating, and a typed PPR traversal with fixed damping and type mixture. For information extraction, a single GPT-4.1

Table 2: QA performance (F1 scores) on RAG benchmarks. In our experiments, we use GPT-4.1 mini as the QA reader. Bold values indicate the best scores.

Method	**Simple QA**		**Multi-hop QA**			**Avg**
	NQ	PopQA	2Wiki	HotpotQA	MuSiQue	
Simple Baselines						
Contriever Izacard et al. (2021)	58.9	53.1	41.9	62.3	31.3	49.5
GTR (T5-base) Muennighoff et al. (2024)	59.9	56.2	52.8	62.8	34.6	53.2
Large Embedding Models						
GritLM-7B Ni et al. (2021)	61.3	55.8	60.6	73.3	44.8	59.1
NV-Embed-v2 (7B) Lee et al. (2024)	61.9	55.7	61.5	75.3	45.7	60.0
Structure-Augmented RAG						
RAPTOR Sarthi et al. (2024)	50.7	56.2	52.1	69.5	28.9	51.4
GraphRAG Edge et al. (2024)	46.9	48.1	58.6	68.6	38.5	52.1
LightRAG Guo et al. (2024)	16.6	2.4	11.6	2.4	1.6	6.9
HippoRAG Jimenez Gutierrez et al. (2024)	55.3	55.9	71.8	63.5	35.1	56.3
HippoRAG 2 Gutiérrez et al. (2025)	63.3	56.2	71.0	75.5	48.6	62.9
Policy-RAG	**68.1**	**74.1**	**78.9**	**80.7**	**55.9**	**71.5**

Table 3: Retrieval performance (passage recall@5) on simple QA and multi-hop QA datasets. The compared structure-augmented RAG methods are reproduced with the same LLM and retriever as ours for a fair comparison. GraphRAG and LightRAG are not presented because they do not directly produce passage retrieval results.

Method	**Simple QA**		**Multi-hop QA**			**Avg**
	NQ	PopQA	2Wiki	HotpotQA	MuSiQue	
Simple Baselines						
Contriever Izacard et al. (2021)	54.6	43.2	57.5	75.3	46.6	55.4
GTR (T5-base) Muennighoff et al. (2024)	63.4	49.4	67.9	73.9	49.1	60.7
Large Embedding Models						
GritLM-7B Ni et al. (2021)	76.6	50.1	76.0	92.4	65.9	72.2
NV-Embed-v2 (7B) Lee et al. (2024)	75.4	51.0	76.5	94.5	69.7	73.4
Structure-Augmented RAG						
RAPTOR Sarthi et al. (2024)	68.3	48.7	66.2	86.9	57.8	65.6
HippoRAG Jimenez Gutierrez et al. (2024)	44.4	53.8	90.4	77.3	53.2	63.8
HippoRAG 2 Gutiérrez et al. (2025)	78.0	51.7	90.4	96.3	74.7	78.2
Policy-RAG	**82.0**	**83.4**	**98.1**	**97.1**	**78.9**	**87.9**

mini model is used for named-entity recognition for similarity and retrieval scoring we use the text-embedding-3-large model OpenAI (2025; 2024). Around the seeded entities, candidate facts are ranked by the embedding retriever, and the top-5 triples are retained for filtering. The QA module conditions on the top-5 retrieved passages as evidence, using an evidence-first prompt and deterministic decoding. Entity scores are projected to passages with light length specificity normalization and combined with a transparent additive re-ranking. All fixed weights, thresholds, and run-time settings are provided in Appendix E.

5 Results and Discussions

We evaluate on three established multi-hop QA benchmarks HotpotQA, 2Wiki, and MuSiQue using a single decoding engine (GPT-4.1 mini) conditioned strictly on evidence retrieved by the policy-driven graph controller Yang et al. (2018); Ho et al. (2020); Trivedi et al. (2022). Answer quality is measured with token-level F1 and EM retrieval quality is assessed with recall@k, MRR, and nDCG, along with diagnostics that probe evidence sufficiency and robustness. Summary results appear in Table 2 (QA performance) and Table 3 (retrieval performance).

Overall performance. PolicyRAG attains strong accuracy across all benchmarks 80.7 F1 on HotpotQA, 78.9 F1 on 2Wiki, and 55.9 F1 on MuSiQue while remaining training-free and fully auditable. Retrieval is both early and precise recall@5 reaches 97.1 on HotpotQA, 98.1 on 2Wiki, and 78.9 on MuSiQue, indicating that gold evidence is surfaced with a small candidate budget. On MuSiQue's distractor-heavy setting, the keep–drop controller maintains competitive F1 and preserves early-rank coverage, suggesting the policy's bias toward compact, high-yield evidence sets rather than broader context stuffing. We also validate transfer on domain-specific corpora (Legal, Health, History) without retraining details appear in Appendix C.

Ranking behaviour and coverage. Ranked lists concentrate verifiable evidence near the top. MRR and nDCG consistently exceed dense-retriever and reader-only controls, reflecting more stable ordering under matched compute. On 2Wiki, HopsRecovered@k saturates quickly, consistent with efficient discovery of both ends of the two-hop chain HotpotQA shows a similar pattern on bridge questions. answer-supported@k tracks recall@k closely, implying that answers are typically derivable from retrieved passages without relying on ungrounded model priors.

Ablation study. We ablate the controller along four axes seeding, gating, connectivity, and scoring to isolate which decisions drive end-to-end gains. Removing seed policy specificity yields noisier traversals and weaker early-rank recall. Disabling local fact gating expands the candidate set but depresses answer-supported@k and harms MRR and nDCG, indicating that compact neighborhoods, not larger ones, improve answerability detailed outcomes appear in Table 4. Dropping synonymy edges reduces robustness to aliasing, while removing typed relation edges weakens multi-hop discovery in both cases, early-rank coverage declines even when recall at larger k is similar, underscoring the role of typed connectivity in arriving early. Turning off length and specificity normalization biases projection toward long or frequent passages and degrades ordering quality, and omitting the transparent re-ranking flattens the head of the list by underweighting title and alias hits as well as multi-seed coverage. Adjusting the PPR damping factor exhibits the expected precision exploration trade-off lower damping explores more but diffuses into off-path regions, whereas higher damping is conservative but risks under-exploration.

Table 4: Ablation study: effect of graph design choices on passage recall@5 in multi-hop QA.

Method	2Wiki	HotpotQA	MuSiQue	Avg
PolicyRAG	98.1	97.1	78.9	91.3
w/ NER to node	98.9	79.6	58.0	78.8
w/ Query to node	73.2	69.1	49.1	63.8
w/o Passage Node	98.0	89.7	67.9	85.2
w/o Filter	98.4	96.2	77.2	90.6

Efficiency and interpretability. At matched token and latency budgets, the controller processes a similar number of passages per query as baselines but surfaces fewer irrelevant items, yielding smaller, easier-to-audit evidence sets. Per-query traces seeds, kept facts, paths, and scores make the retrieval path auditable end to end and immediately actionable for policy edits, enabling governance without retraining while sidestepping the overheads of long-context expansion. Overall, the findings show that a compact policy over a symbolic graph delivers accurate, efficient, and interpretable multi-hop reasoning, and remains robust in distractor-heavy settings.

6 Conclusion

PolicyRAG reframes the prompt as a compact retrieval policy over a symbolic graph, pairing targeted seeding, precise fact gating, and typed traversal to surface, compact verifiable evidence sets. Building on these results, we are currently investigating path-based retrieval policies that traverse longer relational chains while preserving efficiency and per-query traceability. Our near-term focus is to tighten the coupling between gating and traversal to further lift F1. We are also broadening evaluation to large, domain-specific corpora, moving beyond preliminary probes toward rigorously designed, systematic studies. Together, these efforts will advance PolicyRAG toward a path-level retrieval paradigm that preserves accuracy at scale while remaining editable, auditable, and practical.

REFERENCES

Jacob Andreas, Marcus Rohrbach, Trevor Darrell, and Dan Klein. Neural module networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 39–48, 2016.

Akari Asai, Kazuma Hashimoto, Hannaneh Hajishirzi, Richard Socher, and Caiming Xiong. Learning to retrieve reasoning paths over wikipedia graph for question answering. In *Proceedings of the 8th International Conference on Learning Representations*, 2020.

Akari Asai, Xinying Wu, Ruiqi Zhang, Hannaneh Hajishirzi, Mike Lewis, Wen-tau Yih, and Yejin Choi. Self-rag: Learning to retrieve, generate, and critique for language modeling. *arXiv preprint arXiv:2310.11511*, 2023.

Shehzaad Dhuliawala, Mojtaba Komeili, Jing Xu, Roberta Raileanu, Xian Li, Asli Celikyilmaz, and Jason Weston. Chain-of-verification reduces hallucination in large language models. *arXiv preprint arXiv:2309.11495*, 2023.

Darren Edge, Ha Trinh, Newman Cheng, Joshua Bradley, Alex Chao, Apurva Mody, Steven Truitt, and Jonathan Larson. From local to global: A graph rag approach to query-focused summarization. *arXiv preprint arXiv:2404.16130*, 2024.

Zirui Guo, Lianghao Cheng, Yangsibo Ren, Hongye Gao, Cheng Ma, Jiaheng Zhang, Zheng Liu, Jinyang Lin, and Yuan Yao. Lightrag: Simple and fast retrieval-augmented generation. *arXiv preprint arXiv:2410.05779*, 2024.

Bernal Jiménez Gutiérrez, N McNeal, C Washington, Y Chen, L Li, H Sun, and Y Su. Thinking like transformers: Dynamic knowledge graphs for multi-hop reasoning. In *Advances in Neural Information Processing Systems*, volume 37, pp. 12847–12862, 2024.

Bernal Jiménez Gutiérrez, N McNeal, C Washington, Y Chen, L Li, H Sun, and Y Su. Hipporag 2: Enhanced retrieval-augmented generation with neurobiological memory principles. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics*, 2025. forthcoming.

Taher H Haveliwala. Topic-sensitive pagerank. In *Proceedings of the 11th international conference on World Wide Web*, pp. 517–526, 2002.

Xanh Vo Ho, Anh-Minh-Hieu Nguyen, Saku Sugawara, and Akiko Aizawa. Constructing a multi-hop qa dataset for comprehensive evaluation of reasoning steps. In *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 3767–3781, 2020.

Gautier Izacard, Mathilde Caron, Lucas Hosseini, Sebastian Riedel, Piotr Bojanowski, Armand Joulin, and Edouard Grave. Unsupervised dense information retrieval with contrastive learning. *arXiv preprint arXiv:2112.09118*, 2021.

Glen Jeh and Jennifer Widom. Scaling personalized web search. *Proceedings of the 12th International Conference on World Wide Web*, pp. 271–279, 2003.

Bernardo Jimenez Gutierrez, Yixuan Shu, Yifan Gu, Michihiro Yasunaga, and Yixin Su. Hipporag: Neurobiologically inspired long-term memory for large language models. In *Advances in Neural Information Processing Systems*, volume 37, pp. 59532–59569, 2024.

Mingchen Jin, Haochen Wang, Yile Zhang, Kai Liu, and Jun Zhao. Graph chain-of-thought: Augmenting large language models by reasoning on graphs. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics*, volume 1, pp. 4532–4547, 2024.

Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. Dense passage retrieval for open-domain question answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pp. 6769–6781, 2020.

Chen Lee, Rajat Roy, Ming Xu, Jonathan Raiman, Mohammad Shoeybi, Bryan Catanzaro, and Wei Ping. Nv-embed: Improved techniques for training llms as generalist embedding models. *arXiv preprint arXiv:2405.17428*, 2024.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020a.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33: 9459–9474, 2020b.

Zehan Li, Xin Zhang, Yanzhao Zhang, Dingkun Long, Pengjun Xie, and Meishan Zhang. Towards general text embeddings with multi-stage contrastive learning. *arXiv preprint arXiv:2308.03281*, 2023.

Alexander Mallen, Akari Asai, Victor Zhong, Rajarshi Das, Daniel Khashabi, and Hannaneh Hajishirzi. When not to trust language models: Investigating effectiveness of parametric and non-parametric memories. *arXiv preprint arXiv:2212.10511*, 2022.

Niklas Muennighoff, Hao Su, Liang Wang, Nan Yang, Furu Wei, Tianle Yu, Amanpreet Singh, and Douwe Kiela. Generative representational instruction tuning. In *The Thirteenth International Conference on Learning Representations (ICLR)*, 2024.

Jianmo Ni, Chen Qu, Jing Lu, Zhuyun Dai, Gustavo Hernandez Abrego, Ji Ma, Vincent Y. Zhao, Yi Luan, Keith B. Hall, Ming-Wei Chang, et al. Large dual encoders are generalizable retrievers. *arXiv preprint arXiv:2112.07899*, 2021.

Rodrigo Nogueira and Kyunghyun Cho. Passage re-ranking with bert. *arXiv preprint arXiv:1901.04085*, 2019.

OpenAI. text-embedding-3-large: Model overview, 2024. Model documentation; accessed September 23, 2025.

OpenAI. Gpt-4.1 mini: Model overview, 2025. Model documentation; accessed September 23, 2025.

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. *Stanford InfoLab*, 1999.

Boci Peng, Yun Zhu, Yongchao Liu, Xiaohe Bo, Haizhou Shi, Chuntao Hong, Yan Zhang, and Siliang Tang. Graph retrieval-augmented generation: A survey. *arXiv preprint arXiv:2408.08921*, 2024.

Stephen E Robertson and Steve Walker. Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In *SIGIR'94: Proceedings of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, organised by Dublin City University*, pp. 232–241. Springer, 1994.

Parth Sarthi, Salman Abdullah, Aditi Tuli, Shubh Khanna, Anna Goldie, and Christopher D Manning. Raptor: Recursive abstractive processing for tree-organized retrieval. *arXiv preprint arXiv:2401.18059*, 2024.

Haitian Sun, Bhuwan Dhingra, Manzil Zaheer, Kathryn Mazaitis, Ruslan Salakhutdinov, and William Cohen. Open domain question answering using early fusion of knowledge bases and text. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 4231–4242, 2018.

Jiashuo Sun, Chengjin Xu, Lumingyuan Tang, Saizhuo Wang, Chen Lin, Yeyun Gong, Nan Duan, and Ming Zhou. Think-on-graph: Deep and responsible reasoning of large language model on knowledge graph. In *Proceedings of the 12th International Conference on Learning Representations*, 2024.

Nandan Thakur, Nils Reimers, Andreas Rücklé, Abhishek Srivastava, and Iryna Gurevych. Beir: A heterogenous benchmark for zero-shot evaluation of information retrieval models. *arXiv preprint arXiv:2104.08663*, 2021.

Harsh Trivedi, Niranjan Balasubramanian, Tushar Khot, and Ashish Sabharwal. Musique: Multihop questions via single-hop question composition. *Transactions of the Association for Computational Linguistics*, 10:539–554, 2022.

Yuxuan Wang, Ruijia Ren, Jian Li, Wayne Xin Zhao, Jing Liu, and Ji-Rong Wen. Rear: A relevance-aware retrieval-augmented framework for open-domain question answering. *arXiv preprint arXiv:2402.17497*, 2024.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35:24824–24837, 2022.

Jason Weston, Sumit Chopra, and Antoine Bordes. Memory networks. *arXiv preprint arXiv:1410.3916*, 2014.

Lihui Xu, Wenwu Zhang, Hongru Chen, Qian Liu, and Jian Tang. Generate-on-graph: Treat llm as both agent and kg in incomplete knowledge graph question answering. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 8542–8557, 2024.

Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William W Cohen, Ruslan Salakhutdinov, and Christopher D Manning. Hotpotqa: A dataset for diverse, explainable multi-hop question answering. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 2369–2380, 2018.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik R Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*, 2022.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. *Advances in Neural Information Processing Systems*, 36, 2023.

Michihiro Yasunaga, Hongyu Ren, Antoine Bosselut, Percy Liang, and Jure Leskovec. Qa-gnn: Reasoning with language models and knowledge graphs for question answering. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, volume 1, pp. 535–546, 2021.

Xikun Zhang, Antoine Bosselut, Michihiro Yasunaga, Hongyu Ren, Percy Liang, Christopher D Manning, and Jure Leskovec. Greaselm: Graph reasoning enhanced language models. In *Proceedings of the 10th International Conference on Learning Representations*, 2022a.

Zhuosheng Zhang, Aston Zhang, Mu Li, and Alex Smola. Automatic chain of thought prompting in large language models. *arXiv preprint arXiv:2210.03493*, 2022b.

APPENDIX

This appendix provides an in-depth overview of the components underlying our approach, with detailed elaboration on the following key aspects:

- Appendix A: LLM Prompts Policy
- Appendix B: PolicyRAG Pipeline walk-through
- Appendix C: Detailed Experimental Results
- Appendix D: Error Analysis
- Appendix E: Implementation Details and Hyperparameters

A LLM PROMPT POLICIES FOR POLICYRAG

We provide an example demonstration of the fact filtering prompt policy in Figure 4. Alongside this, we define harmonized prompt policies for entity extraction, triple construction, seed selection, domain adaptation, and answer generation together, these policies deliver strong and consistent results across our evaluations.

Fact Filtering Prompt Policy

Instruction:

You are the fact-filtering gate in PolicyRAG, a high-stakes QA system. Your job is to select ONLY the facts that are necessary and sufficient to answer (or strongly support answering) the query, including via short multi-hop chains.

You must select up to 4 relevant facts from the provided candidate list that have a strong connection to the query, aiding reasoning and supporting an accurate answer.

Policy:
1) Direct answer triples > contextual ones
2) Minimal bridges that connect query → target fact
3) Specific/temporal/relational > generic
4) Exact entity string > alias
5) Remove near-duplicates; smallest sufficient set

The output must be in strict JSON format, e.g. {"fact": [["s1","p1","o1"], ["s2","p2","o2"]]}
If no facts are relevant, return: {"fact": []}

You must only use facts from the candidate list and never generate new facts. Accuracy is critical because your filtered facts directly guide the reasoning process in this system.

Demonstration:

Question: When was the person who Messi's goals in Copa del Rey compared to get signed by Barcelona?
Fact Before Filter: {"fact":[
["barcelona","won","semi-finals first leg against getafe 5–2"],
["messi","scored","goal bringing comparison to diego maradona's goal of the century"],
["barcelona","lost","second leg 4–0"],
["diego maradona","was signed by barcelona","june 1982"],
["maradona","signed from","boca juniors"]
]}
Fact After Filter: {"fact":[
["messi","scored","goal bringing comparison to diego maradona's goal of the century"],
["diego maradona","was signed by barcelona","june 1982"]
]}

Figure 4: LLM prompts policy for fact filtering

B POLICYRAG PIPELINE WALK-THROUGH

Table 5 shows the end-to-end online loop parsing a question into candidate entities and local facts, keep–drop gating to remove distractors, typed PPR over the entity graph, projection to passages with light normalization, transparent re-ranking, top-k evidence selection, and an emitted per-query trace for audit and reuse.

Table 5: PolicyRAG query processing and response generation. Compact view of entities, facts, and reasoning steps leading to the final answer.

Original Query	What county contains the city with a radio station that broadcasts to the capital city of the state where the Peace Center is located?
Gold Answer	Richland County
Entities	Peace Center; Greenville; South Carolina; WWNQ; Forest Acres; WFFG-FM; Warren County
Facts (Triples)	(Peace Center, located in, Greenville) (Greenville, in, South Carolina) (WWNQ, licensed to, Forest Acres) (WFFG-FM, located in, Warren County)
Retrieved Info	Forest Acres is in Richland County, SC; WWNQ serves Columbia (capital of SC); Peace Center is in Greenville, SC
Response	Forest Acres is a city in Richland County, South Carolina, United States.
Reasoning Chain	Peace Center → Greenville → South Carolina → Columbia (capital) → WWNQ → Forest Acres → Richland County

C DETAILED EXPERIMENTAL RESULTS

We report full QA and retrieval metrics for HotpotQA, 2Wiki, and MuSiQue under matched-compute settings; summaries appear in Table 6 and Table 7. We also evaluate domain-specific corpora (Legal, Health, History) without retraining, with illustrative cases in Figure 5. In QA results (GPT-4.1 mini as reader), PolicyRAG attains strong EM and F1 with clear gains on MuSiQue and 2Wiki. Retrieval quality shows higher recall@2 and recall@5, indicating earlier coverage of gold evidence compared to competitive retrievers.

Legal	Health	History
Q1: When was the Restructuring Support Agreement signed? **Human Ans:** July 5, 2020 **PolicyRAG Ans:** Executed on July 5, 2020	Q1: Which country's traditional diet is linked to lower heart disease? **Human Ans:** Crete **PolicyRAG Ans:** Crete, Greece (Mediterranean diet) linked to low heart disease	Q1: In which year did the first major evacuation in Britain take place? **Human Ans:** 1939 **PolicyRAG Ans:** September 1939, outbreak of WWII, ~2M civilians evacuated
Q2: Who is the debtor company under this agreement? **Human Ans:** Endologix, Inc. **PolicyRAG Ans:** The debtor is Endologix, Inc., a Delaware corporation	Q2: Which vitamins are absorbed only with fats? **Human Ans**: A, D, E, K **PolicyRAG Ans:** Vitamins A, D, E, and K are fat-soluble, require dietary fats	Q2: What important social report was influenced by the poverty revealed during evacuation? **Human Ans:** Beveridge Report **PolicyRAG Ans:** Poverty exposure influenced the Beveridge Report (1942), foundation of welfare state
Q3: What does the term "Alternative Transaction" include? **Human Ans:** Merger, refinancing **PolicyRAG Ans:** Includes refinancing, recapitalization, merger, acquisition, or business combination	Q3: Where is the "Cold Spot" for diabetes mentioned in the book? **Human Ans:** Copper Canyon **PolicyRAG Ans:** Copper Canyon, Mexico; Tarahumara diet linked to low diabetes	Q3: What essential item were schoolchildren required to carry during evacuation? **Human Ans:** Gas mask **PolicyRAG Ans:** Each child carried a gas mask as protection against gas attacks

Figure 5: Human-verified domain examples (Legal, Health, History). PolicyRAG surfaces compact evidence and grounded answers without retraining.

Table 6: QA performance EM / F1 scores on RAG benchmarks.

Method	Simple QA		Multi-hop QA			Avg
	NQ	PopQA	2Wiki	HotpotQA	MuSiQue	
Simple Baselines						
Contriever Izacard et al. (2021)	45.0 / 58.9	46.1 / 53.1	38.1 / 41.9	51.3 / 62.3	24.1 / 31.3	33.0 / 49.5
GTR (T5-base) Muennighoff et al. (2024)	45.5 / 59.9	43.2 / 56.2	49.2 / 52.8	50.6 / 62.8	25.8 / 34.6	34.8 / 53.2
Large Embedding Models						
GritLM-7B Ni et al. (2021)	46.8 / 61.3	42.8 / 55.8	55.8 / 60.6	60.7 / 73.3	33.6 / 44.8	41.3 / 59.1
NV-Embed-v2 (7B) Lee et al. (2024)	43.5 / 59.9	41.7 / 51.5	54.4 / 60.8	57.3 / 71.0	32.8 / 46.0	41.8 / 60.0
Structure-Augmented RAG						
RAPTOR Sarthi et al. (2024)	37.8 / 54.5	41.9 / 55.1	39.7 / 48.4	50.6 / 64.7	27.7 / 39.2	36.9 / 49.7
GraphRAG Edge et al. (2024)	38.0 / 55.5	30.7 / 51.3	45.7 / 61.0	51.4 / 67.6	27.0 / 42.0	36.0 / 52.6
LightRAG Guo et al. (2024)	2.8 / 15.4	1.9 / 14.8	2.5 / 12.1	9.9 / 20.2	2.0 / 9.3	3.6 / 13.9
HippoRAG Jimenez Gutierrez et al. (2024)	37.2 / 52.2	42.5 / 56.2	59.4 / 67.3	46.3 / 60.0	24.0 / 35.9	38.9 / 51.2
HippoRAG 2 Gutiérrez et al. (2025)	43.4 / 60.0	41.7 / 55.7	60.5 / 69.7	56.3 / 71.1	35.0 / 49.3	44.3 / 58.1
Policy-RAG	**64.2 / 68.1**	**71.0 / 74.1**	**76.4 / 78.9**	**79.9 / 80.7**	**51.2 / 55.9**	**68.54 / 71.5**

Table 7: Retrieval performance passage recall@2 / recall@5 on simple QA and multi-hop QA datasets.

Method	Simple QA		Multi-hop QA			Avg
	NQ	PopQA	2Wiki	HotpotQA	MuSiQue	
Simple Baselines						
Contriever Izacard et al. (2021)	29.1 / 54.6	27.0 / 43.2	46.6 / 57.5	58.4 / 75.3	34.8 / 46.6	39.2 / 55.4
GTR (T5-base) Muennighoff et al. (2024)	35.0 / 63.4	40.1 / 49.4	60.2 / 67.9	59.3 / 73.9	37.4 / 49.1	46.4 / 60.7
Large Embedding Models						
GTE-Qwen2-7B-Instruct Ni et al. (2021)	44.7 / 74.3	47.7 / 50.6	66.7 / 74.8	75.8 / 89.1	48.1 / 63.6	56.6 / 70.5
NV-Embed-v2 (7B) Lee et al. (2024)	45.3 / 75.4	45.3 / 51.0	67.1 / 76.5	84.1 / 94.5	52.7 / 69.7	58.9 / 73.4
Structure-Augmented RAG						
RAPTOR (GPT-4o-mini) Sarthi et al. (2024)	40.5 / 69.4	37.2 / 48.1	58.4 / 66.0	78.6 / 90.2	49.1 / 61.0	52.8 / 67.0
HippoRAG Jimenez Gutierrez et al. (2024)	21.6 / 45.1	36.5 / 52.2	68.4 / 87.0	60.1 / 78.5	41.8 / 52.4	45.7 / 63.0
HippoRAG 2 (GPT-4o-mini)	44.4 / 76.4	43.5 / 52.2	74.6 / 90.2	80.5 / 95.7	53.5 / 74.2	59.3 / 77.7
Policy-RAG	**77.0 / 82.0**	**75.1 / 83.4**	**78.9 / 98.1**	**81.5 / 97.1**	**76.3 / 78.9**	**77.7 / 87.9**

D Error Analysis

We find four recurring failure types: alias spillover, ambiguous seeding, hub bias, and long-passage dilution. Small, auditable prompt-policy edits tighter seeding instructions, calibrated type-mix weights, light degree-aware penalties, and stronger length IPF normalization consistently fix these cases without retraining, improving early-rank coverage and answer-supported@k (see Table 8).

Table 8: Representative failures and one-line policy edits. Each edit is a deterministic controller change.

Dataset	Error type	Symptom (trace excerpt)	Policy edit (one line)	Outcome
2Wiki	Ambiguous seeding	Seeds include `Washington` (person) and `Washington` (state); traversal oscillates	Increase seed-specificity prior for person NER tags; require alias_ok on toponyms only if question mentions location	+7% HS@5, earlier arrival on correct entity pair
HotpotQA	Alias spillover	Synonym edge pulls `Mercury` (element) from `Mercury` (planet) context	Reduce η_s in T_E when query contains relation phrases (e.g., "orbited by")	+5 nDCG@10, reduced off-topic hops
MuSiQue	Hub bias	High-degree category node absorbs mass; gold passage ranks 12th	Add hub penalty in rerank: ϕ_{path} rewards short, type-coherent chains; light degree prior	+8 MRR@10, gold rises to top-5
HotpotQA	Long-passage dilution	Very long wiki list outranks concise biographical passage	Strengthen length/IPF normalization: $(\beta, \gamma)\uparrow$	+6 nDCG@10, Answer-Supported@5 aligns with Recall@5

E Implementation Details and Hyperparameters

We initialize retrieval with a compact, auditable policy. Two seed types are used and combined into a single restart distribution for typed PPR.

Seed selection Phrase (entity) seeds come from filtered triples produced by the extraction pass; we keep up to five distinct phrases ranked by the mean score of their surviving triples (embeddings are ℓ_2–normalized before scoring).

Restart distribution. Let $\mathcal{E}_s$ be phrase seeds and $\mathcal{P}_s$ passage seeds. The per-query restart mass over entities is

$$\mathbf{s}(e) \propto \underbrace{\mathrm{align}(e,q)}_{\text{lex/emb match}} \cdot \underbrace{\mathrm{spec}(e)}_{\text{down-weight generic}} \cdot \underbrace{\mathrm{alias\_ok}(e)}_{\{0,1\}},$$

and over passages is $\mathbf{s}(p) \propto w_p \cdot \mathrm{sim}(p,q)$, where w_p balances phrase vs. passage influence. The final restart vector concatenates entity and passage components and is normalized once per query.

Typed PPR and scoring. We diffuse over the entity layer with $T_E = \eta_r A_r + \eta_s A_s$ (relation vs. synonymy), iterate $\mathbf{v}^{(t+1)} = (1-\alpha)T_E^\top \mathbf{v}^{(t)} + \alpha\,\mathbf{s}$ until convergence, then project to passages $\mathbf{u} = B^\top \mathbf{v}^\star$. Light normalization corrects length frequency bias, and a small transparent reranker adds bonuses for title/alias hits, multi-seed coverage, and short, type-coherent paths. We maintain the graph in Memgraph[2] and compute PPR in a read-only setting. Information extraction uses a single GPT-4.1 mini model and similarity retrieval scoring uses text embedder-3-large. Around the seeded entities, we rank nearby facts and retain the top five triples for filtering the QA module conditions on the top five retrieved passages with deterministic decoding.

E.1 Hyperparameters

Unless noted, settings are fixed across datasets. We tune only a small set of knobs on 100 MuSiQue training examples: PPR damping α, type mixture (η_r, η_s), passage/length correction (β, γ), and rerank weights $(\lambda_{\text{title}}, \lambda_{\text{cov}}, \lambda_{\text{path}})$. We cap phrase seeds at 5, facts kept at 5, and QA context at 5 passages. Full values are listed in Table 9.

Table 9: PolicyRAG hyperparameters.

Hyperparameter	**Value**
Synonym threshold	0.7
PPR damping factor	0.5
Generation temperature	0.0

Table 10: GraphRAG vs LightRAG hyperparameters

Parameter	**GraphRAG**	**LightRAG**
Mode	Local	Local
Response	Short	Short
Top-k phrases	60	60
Chunk size	1,200	1,200
Overlap	100	100
Max report len	2,000	–
Max input	8,000	–
Max cluster	10	–
Entity tokens	–	500

[2] https://memgraph.com/docs

E.2 Baselines and Protocol

All methods operate on the identical passage pool under matched token and latency budgets. When public checkpoints and recommended hyperparameters are available, we adopt them; otherwise, we use the reported configurations. Dense retrievers are run with PyTorch and Hugging Face; BM25 uses BM25s. For GraphRAG and LightRAG, we retain the public configurations and ensure compute parity. Full settings appear in Table 10.