Causality Is What You Need

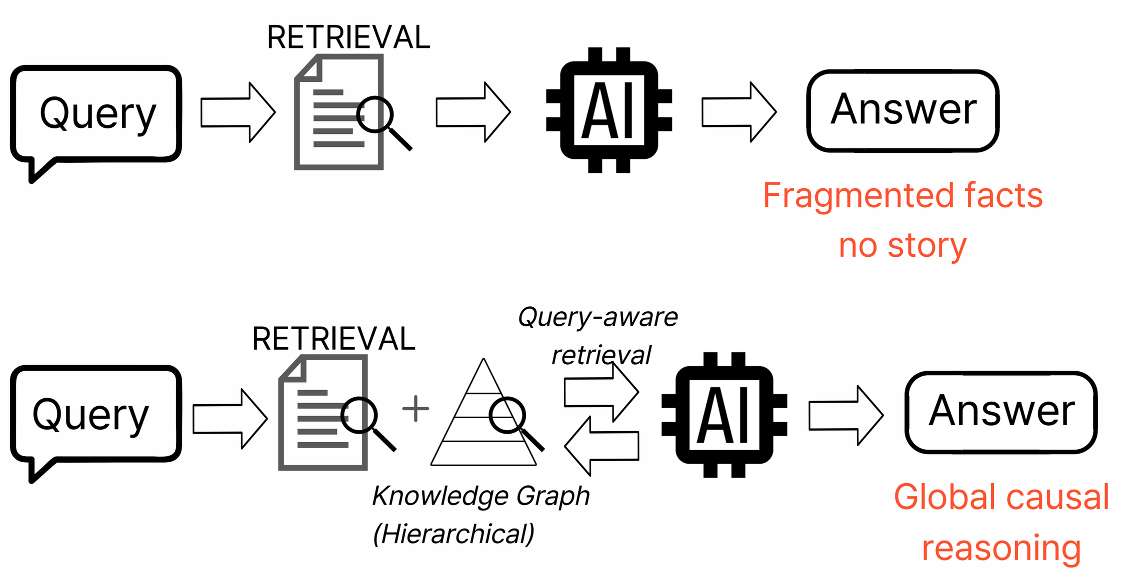
A Novel Hybrid GraphRAG Chatbot Architecture for Explaining Economic Events

HYU DS

Hyundo Jung, Minkyum Kim, Prof. Misuk Kim / Department of Data Science, Hanyang University

ABSTRACT

<Figure : BaselineRAG vs modular hybrid RAG>



Common people (non-economics majors) are exposed every day to **economic phenomena** such as stock price swings, exchange rate changes, and housing market shifts, yet **they rarely grasp the precise causal** mechanisms behind these changes.

Motivation: Developing a chatbot that explains the reasons behind economic phenomena using RAG.

so we built an **embedding-based** RAG chatbot that grounds its **explanations** in a corpus of financial news and introductory **economics**. although **the baseline** worked well for simple factoid questions, it still **struggled with** queries requiring **causal** reasoning and **global integration** of multiple articles.

Limitation: Hyperparameter tuning revealed baseline limits, prompting new architecture exploration.

We therefore **designed** a **hybrid architecture** centered on **GraphRAG, modularizing** based **response algorithms** (e.g., CoRAG, MADAM-RAG, MCTS-RAG, NodeRAG, ReARAG, Typed-RAG) and **routing** each query through an appropriate **pipeline.**
Experiments show that this **modular hybrid RAG** achieves **higher accuracy** when explaining relationships between economic events and their outcomes (e.g., stock price movements) **than the BM25+DPR baseline.**

Conclusion: A new hybrid architecture combining multiple strategies achieves higher accuracy.

TIME LINE

2025.03 ~ 2025.04

- Define task: Define an **economic-explanation chatbot** for real-world phenomena.
- Build dataset: **two corpus** from intro economics texts and financial news.

2025.05 ~ 2025.08

- RAG baseline: **Implemented** a BM25+DPR retriever **RAG.**
- Tuning: Systematically **tuned** chunk sizes, weights, and LangChain RAG **hyperparameters.**

2025.09

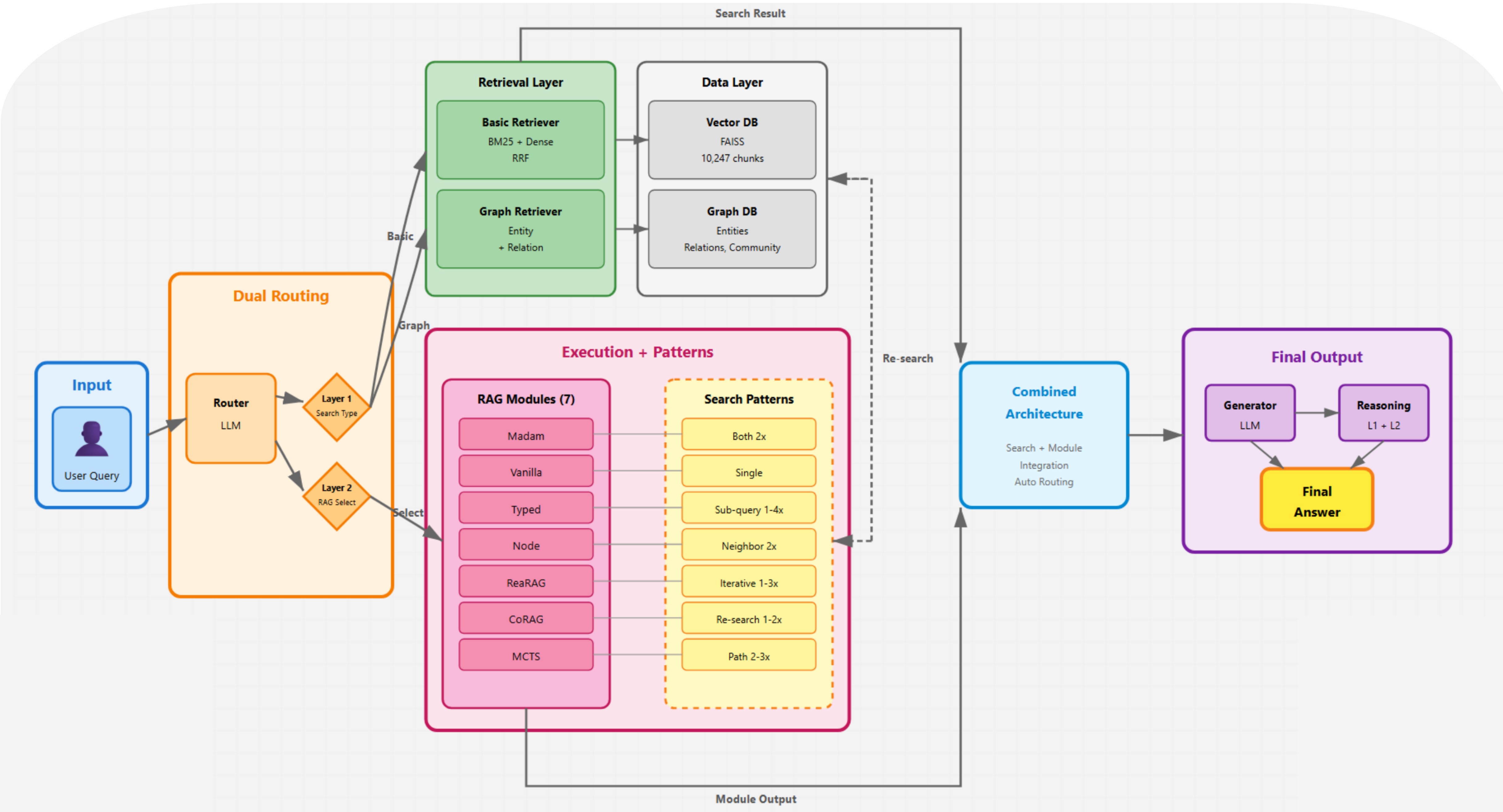
- Baseline RAG limitations: Found **DPR limits** on global, multi-document questions and **surveyed alternative architectures.**

Literature Review

2025.10 ~ 2025.12

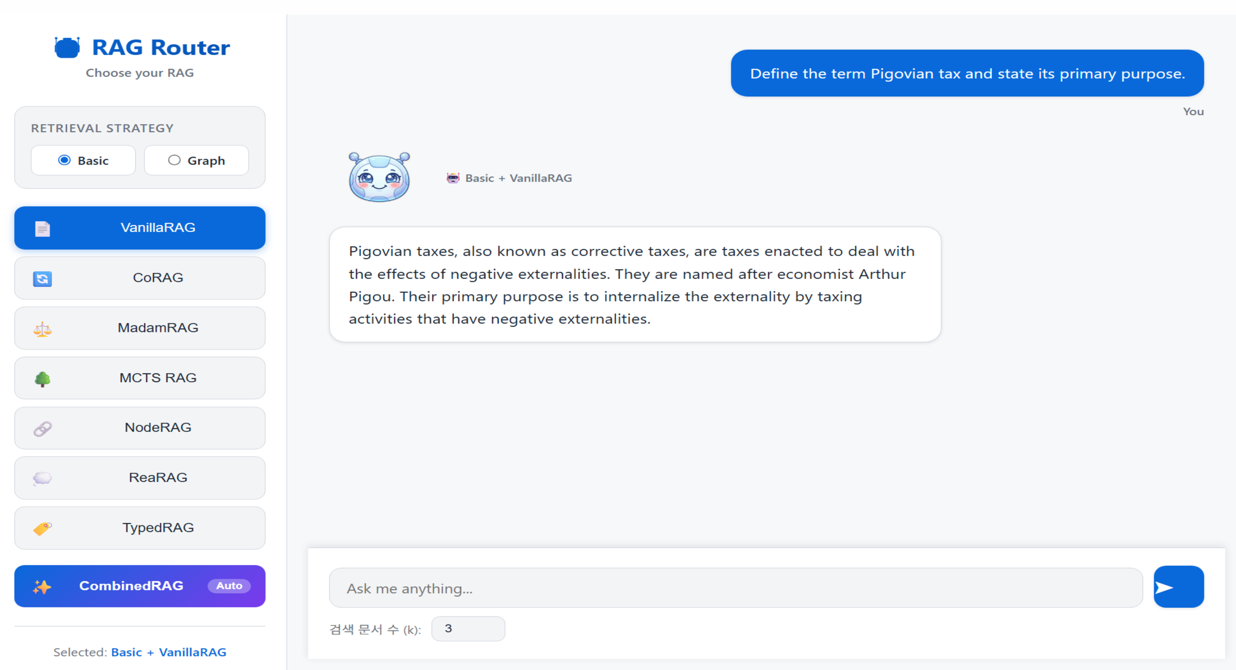
- New hybrid architecture: Built a **modular**, domain-specific **hybrid GraphRAG** pipeline that combines multiple retrievers with graph-structured knowledge.

SYSTEM ARCHITECTURE

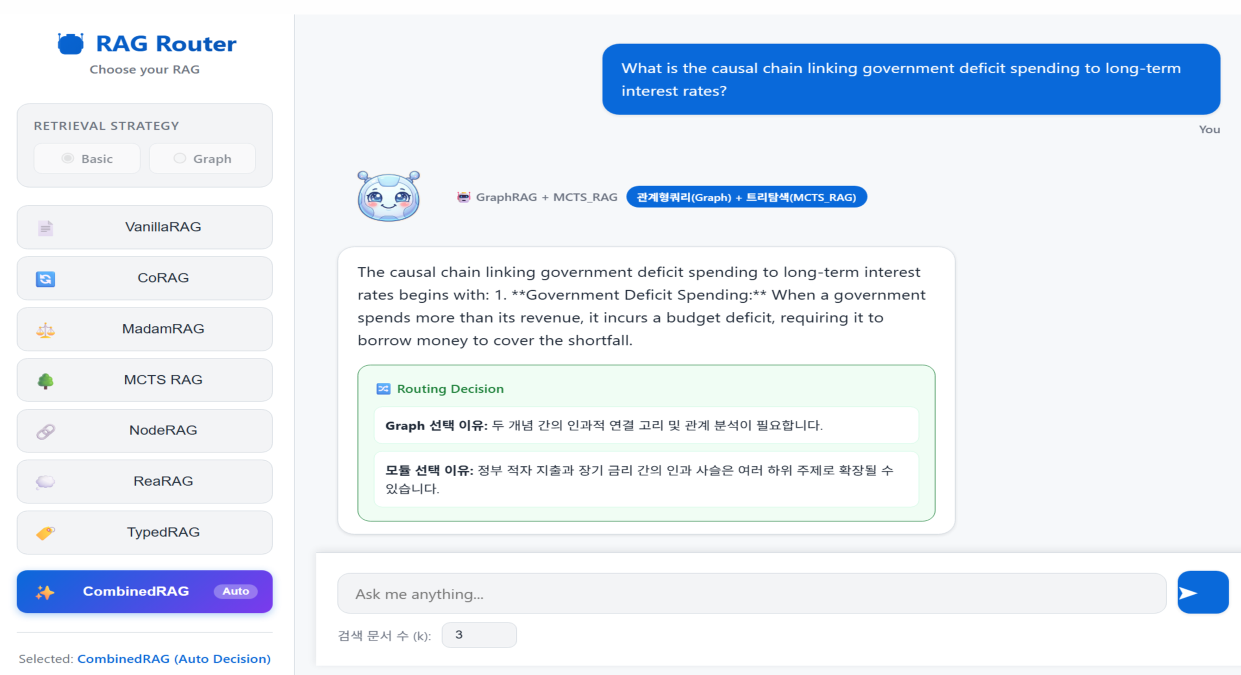


- Dual Routing: **Determines** the retrieval **method** and the corresponding RAG **module** based on the **query type.**
- Layer 1: Selects the appropriate **search method** (either Basic or Graph).
- Layer 2: Selects the **module** optimized for the query from among the **seven** available **RAG architectures.**
- Re-search: Performs **subsequent DB searches** using module-specific **regenerated** queries.
- Reasoning: Provides the **rationale for the selection** made in both Layer 1 and Layer 2.

UI EXAMPLE



User-controlled mode: The user selects a retriever / GraphRAG mode and receives an answer tailored to that strategy.



Automatic routing: The system analyzes the query, selects an **appropriate retrieval / GraphRAG pipeline,** and returns the answer with a brief **rationale.**

PERFORMANCE
Baseline RAG vs Our Hybrid RAG

<Table 1 : Basic RAG modules result >

Architecture	Latency (s)	BERTScore (F1)	Human causal score
VanillaRAG	13.02	0.8667	2
CoRAG	44.82	0.8738	3
MadamRAG	44.60	0.8566	4
MCTS_RAG	54.39	0.8688	3
NodeRAG	30.25	0.8692	5
ReaRAG	45.60	0.8755	4
TypedRAG	28.63	0.8696	3
Basic-Combined (A)	47.60	0.8728	4

<Table 2 : Graph-Augmented RAG modules result >

Architecture	Latency (s)	BERTScore (F1)	Human causal score
Graph + VanillaRAG	108.08	0.8399	4
Graph + CoRAG	171.41	0.8509	4
Graph + NodeRAG	337.32	0.8607	5
Graph + ReaRAG	140.90	0.8505	4
Graph + MCTS_RAG	359.64	0.8572	4
Graph + TypedRAG	135.44	0.8596	4
Graph + MadamRAG	162.41	0.8395	5
CombinedRAG (Full Aut)	68.73	0.8683	4

Human causal score(1-5): *how clearly answer explains cause–effect relationships.*

Query Why did airline stocks go up even though there was news about higher oil prices?

BaselineRAG Higher oil prices increase fuel costs for airlines, which can hurt profits and lead to lower stock prices.
→ Facts only, no real why (low human score)

hybrid RAG Oil prices rose, but very strong travel demand let airlines sell more tickets and raise fares, so investors expected higher profits and the stocks went up.
→ Clear cause–effect story (high human score)

CONCLUSION: Strengths & Limitations

Conclusion

- **Hybrid GraphRAG chatbot** that explains real economic events using textbook economics and financial news.
- Query-based routing and graph reasoning improve **"why" questions** and multi-document answers **over a vanilla RAG baseline.**
- **Applicable to other retrieval-based domains** (e.g., legal or internal policy QA) and especially effective on small domains, delivering high accuracy with low latency.

Strengths

- **Higher answer quality** than a vanilla RAG baseline, especially for questions that **connect multiple events and concepts.**
- **Routing** picks a **fast pipeline** with good accuracy, and graph modules turn raw facts into clear causal stories.

Limitations

- Graph building and graph reasoning add extra cost, so responses can be slow on very large graphs or very complex queries.

Acknowledgments & Contact

This work was carried out under the guidance of Prof. Misuk Kim, whom we sincerely thank for her continuous support and feedback.

Hyundo Jung : oen123456@hanyang.ac.kr
Minkyum Kim: mkk114@hanyang.ac.kr