General Design Description for the Proposed Multi-Agent Retrieval-Augmented Generation (RAG) Framework

Task Definition

The objective of the Multi-Agent RAG Framework is to dynamically integrate and reason over real-time, cross-domain knowledge from heterogeneous data sources. This system aims to address complex, interdisciplinary user queries by synthesizing relevant information from multiple knowledge bases and performing structured reasoning to deliver accurate, context-rich responses.

Motivation

Current knowledge integration systems exhibit several limitations, including static data retrieval, limited reasoning capabilities, and inadequate user interaction mechanisms. The motivation for this proposed design is to overcome these challenges by creating a dynamic, interactive, and comprehensive system capable of real-time knowledge updates, advanced logical reasoning, and responsive user interaction, thereby significantly enhancing knowledge management efficiency and user satisfaction.

Architecture Overview

The proposed framework consists of the following specialized agents, each responsible for distinct aspects of knowledge retrieval, processing, and interaction:

1. **Coordinator Agent**
   - Handles user input, analyzes query intent, and assigns sub-tasks to relevant agents.
   - Aggregates responses from all other agents to provide cohesive output.
2. **Retriever Agent**
   - Employs advanced retrieval methods (e.g., RAG-Fusion) to perform dynamic, multi-source document retrieval from databases such as Arxiv, Wikipedia, and structured experiment data repositories.
3. **Analyst Agent**
   - Extracts structured information using semantic segmentation, named entity recognition, and relation extraction.
   - Constructs knowledge graphs to represent relationships clearly.
4. **Reasoner Agent**
   - Applies advanced reasoning methodologies (Chain-of-Thought and Graph-of-Thought frameworks).
   - Provides logically coherent and traceable reasoning paths.
5. **Validator Agent**
   - Verifies logical consistency and factual accuracy of reasoned responses.
   - Iteratively enhances answer quality through feedback loops.
6. **Interactive Agent**

o   Facilitates sophisticated user interaction through multi-turn dialogue.
o   Incorporates user feedback to refine and optimize system behavior.

Inter-Agent Communication

Agents communicate through standardized protocols (e.g., message queues like Kafka), ensuring asynchronous, efficient interaction and collaboration. Each agent operates as a modular, independently deployable microservice, promoting scalability and maintainability.

Experiments and Evaluation

The system's performance will be assessed through comprehensive evaluations focusing on:

- **Real-Time Response**: Time from query initiation to response delivery.
- **Accuracy and Precision**: Evaluated against manually annotated benchmarks.
- **Logical Consistency and Transparency**: Expert reviews of reasoning paths.
- **User Interaction Quality**: User satisfaction scores obtained through feedback questionnaires.

Innovation

- Introduction of Graph-of-Thought (GoT) methodology for structured, transparent reasoning.
- Incorporation of a dedicated Validator Agent for dynamic result validation.
- Advanced integration of real-time knowledge updates and interactive feedback loops.