

You:

In the balance population simulation process, what factors are related to the birth term and the death term?

Al:

Introduction

In the context of balance population simulation, particularly in systems involving particle aggregation, the dynamics of the population are influenced by two primary terms: the birth term and the death term. These terms are crucial for understanding how the population evolves over time, especially in processes like aggregation where particles combine to form larger entities.

Main Body

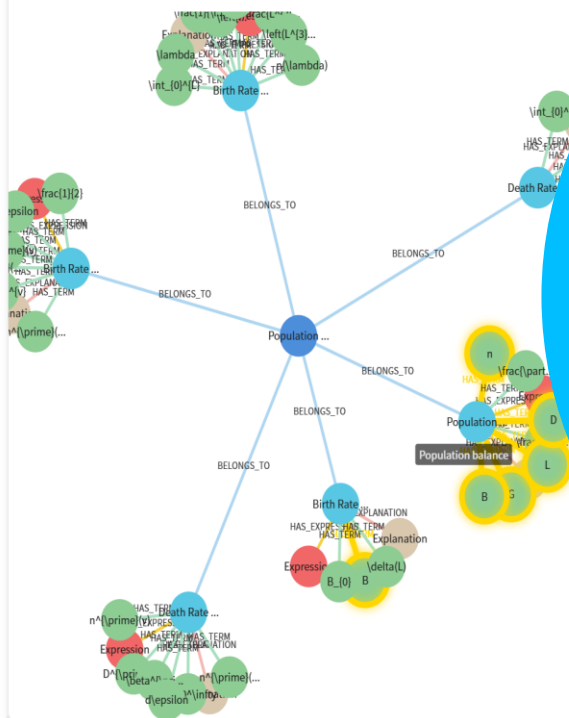
Birth Term

The birth term, denoted as \overline{B}_j , represents the contribution to the j -th moment of the population due to the formation of new particles. In aggregation processes, this typically involves smaller particles combining to form larger particles, which then contribute positively to the population moment. Key factors influencing the birth term include:

- **Particle Size Distribution (PSD):** The distribution of particle sizes at any given time, denoted as $n = n(L, t)$, affects how

Send

Graph Visualization



Scalable GraphRAG-based Knowledge System for Domain Specific Workflow Automation with LLMs

Master thesis 3816

Haonan Chen

Electrical Engineering and
Information Technology

Supervisor: Yuchen Xia



Introduction

Current status of manufacturing industry

- Knowledge-Intensive: complex processes, scientific models
- Management of Knowledge

Traditional knowledge utilization method

- Storage of data in Database
- Large Language Model(LLM) Retrieval

Challenge

How can LLM utilize the local knowledge?

Formula-Centric Knowledge

Knowledge:

Particle Size Balance Model

Formula Expression:

$$\frac{\partial n}{\partial t} + \frac{\partial(Gn)}{\partial L} = B - D$$

Formula Explanation:

This equation is commonly used to describe the dynamic behavior of a particle,...

Explanation of Terms:

- $B=B(L,t)$:The birth rate of the particle,...
- $D=D(L,t)$:The death rate of the particle,...

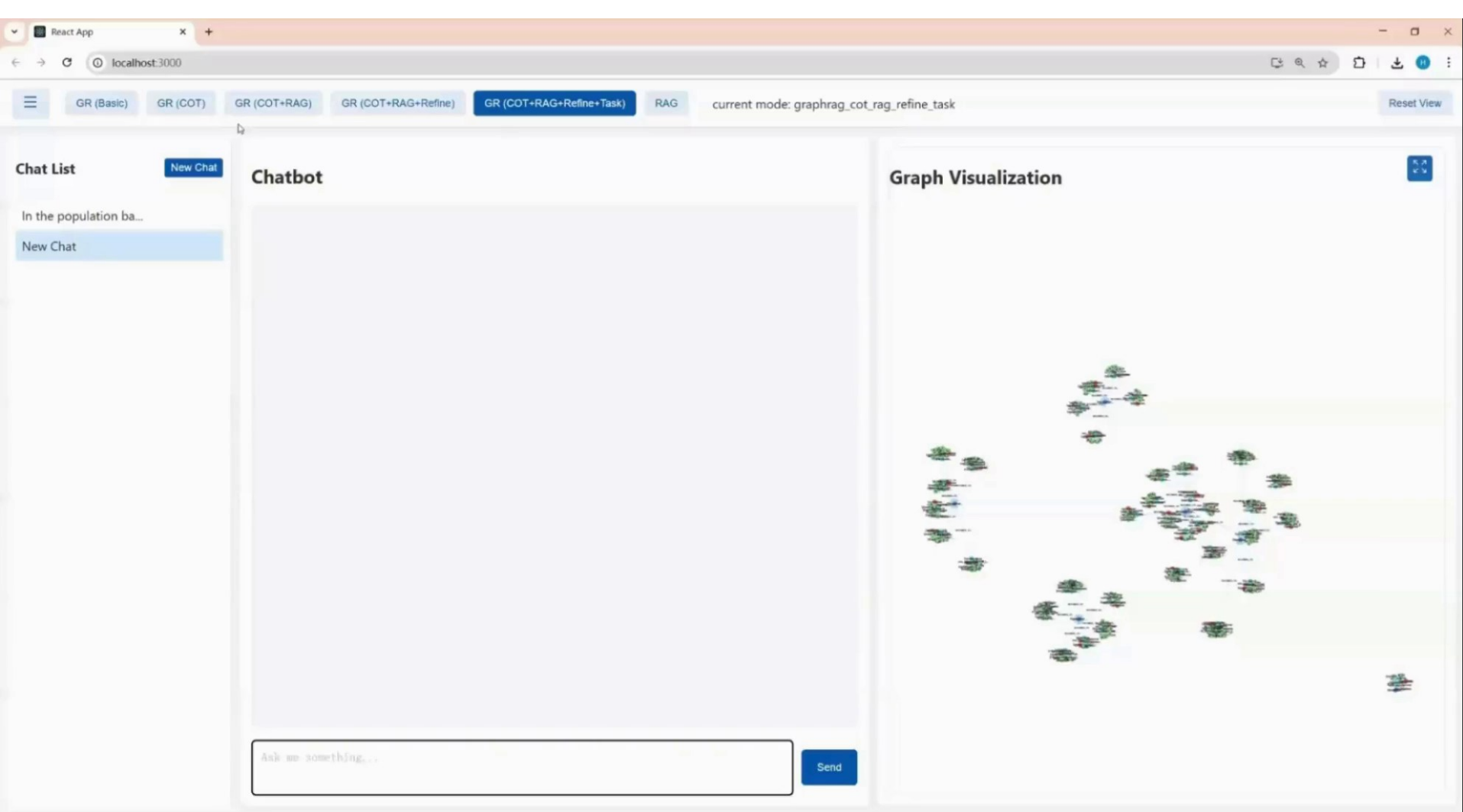
...

Knowledge Source:

Scientific paper

Knowledge System

Knowledge Graph-based Retrieval + LLM Reasoning + QA Chatbot



Contents

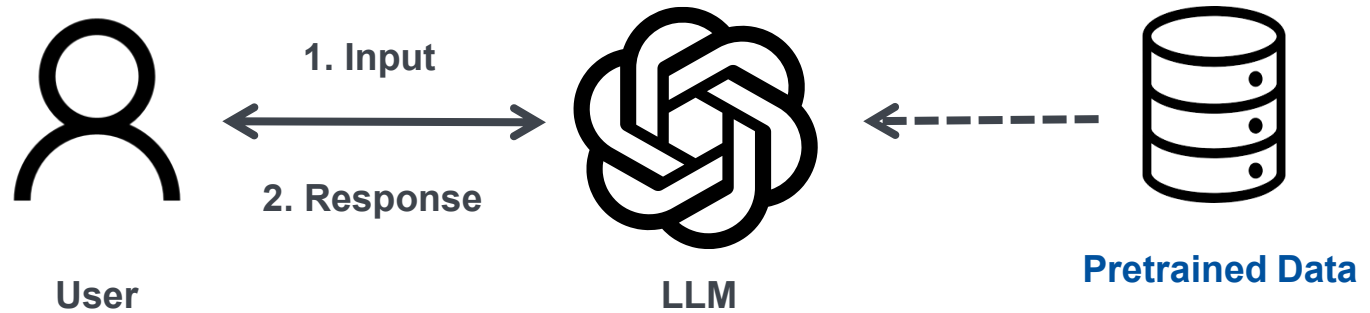
- **Introduction**
- **Background**
- **System design**
- **Evaluation & Analysis**
- **Conclusion & Outlook**

Background

- LLM Only
- Retrieval-Augmented Generation (RAG)
- Graph-based RAG(GraphRAG)

Background

LLM Only



Limitations:

- No Access to Local Knowledge
- Risk of Data Leakage

Question:

How can domain knowledge be incorporated?

Background

Retrieval-Augmented Generation (RAG)

RAG Local Knowledge Utilization

```
Enter your question (or type 'exit' to quit): in the population balance model, what factors are related to the birth term and the death term?
```

```
>>> Retrieving context (RAG retrieval)...
```

```
=== Retrieved Context ===
```

```
--- Retrieved Chunk 1 ---
```

population balance equation models to simulate secondary nucleation processes have been derived. The first simulates attrition as a breakage term, and growth rate is the result of size-dependent solubility. The second model considers attrition as a boundary condition at zero crystal size, where the expression for secondary nucleation rate already takes into account the effect of supersaturation, while the growth rate is size-independent. The two models are proven equivalent in the growth regime, thus

```
--- Retrieved Chunk 2 ---
```

function is $n - \frac{dn}{dL}$. Equation 1 relates the rate of change of number of particles in the differential size range L to $L + dL$, to the rates of growth into and out of that range and the rates of birth and death in the size range.

In writing the population balance as Eq. 1, only growth, of the three phenomena of interest, appears explicitly. Before considering how appropriate birth and death functions might be formed for the other phenomena, a small simplification may be made to

```
--- Retrieved Chunk 3 ---
```

into either the “breakage model” or the “nucleation model”. Model Equations. The model consists of the following population balance equation and solute mass balance:

Limitations:

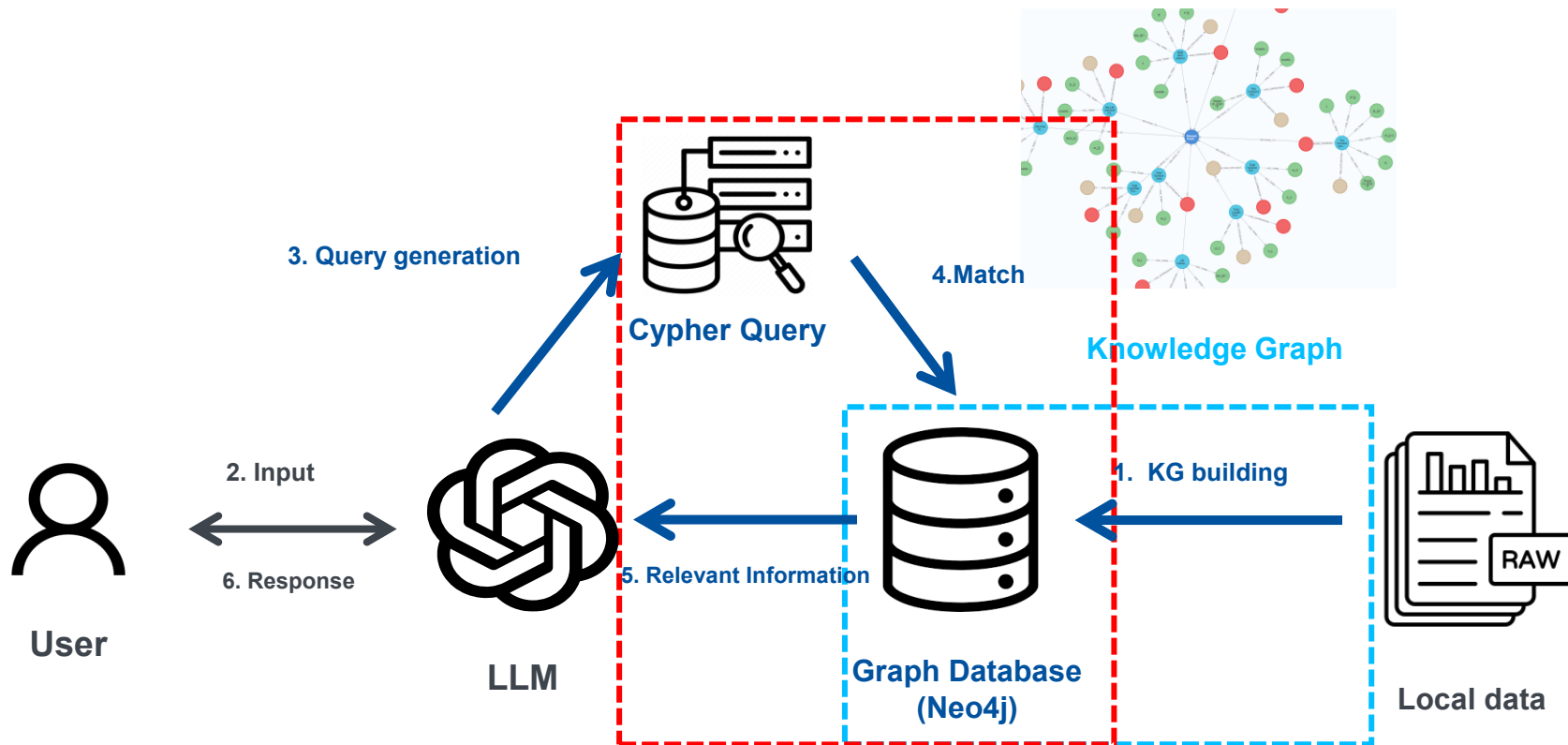
- **Chunk-based Context Loss**
- **Limited Relational Understanding**

Background

Graph-based RAG(GraphRAG)

GraphRAG

- Relationship-driven Reasoning
- Interpretability through Visualization



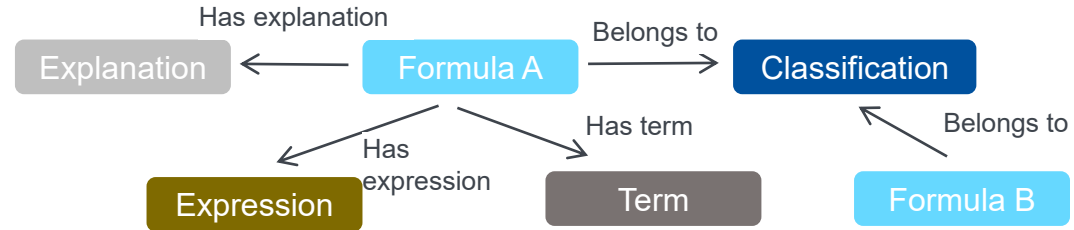
Background

Graph-based RAG(GraphRAG)

User Task

Under the conditions of **aggregation** and **nucleation**, what effects do they have on the **Particle Size Balance Model** ?

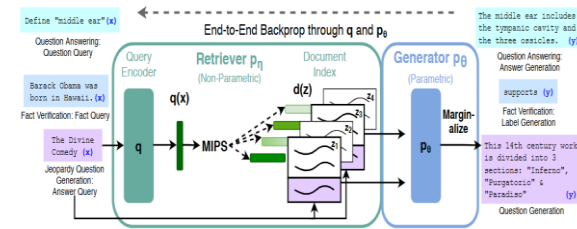
Graph Structure



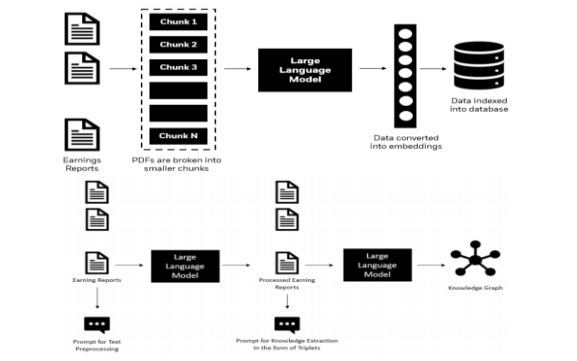
Structured Data

```
{
  Formula Name:[Particle Size Balance Model],
  Formula Explanation:[...aggregation...nucleation],
  Formula Expression:  $\frac{\partial n}{\partial t} + \frac{\partial(Gn)}{\partial L} = B - D$ 
  Formula Source:[...]
  Term:
  • Term Name:[Birth Term], Term Meaning:[...aggregation...nucleation]
  • Term Name:[Death Term], Term Meaning:[...]
  • ...
}
```

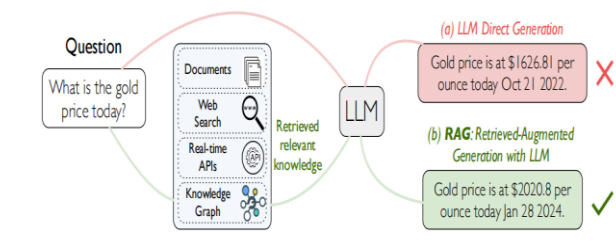
Background Literature Review



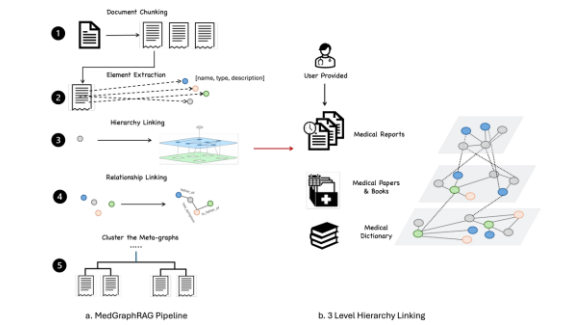
[1] RAG for knowledge-intensive NLP



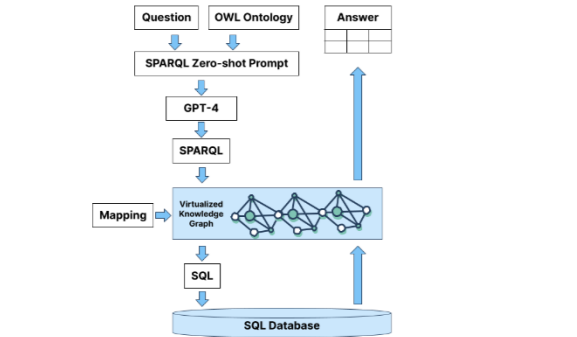
[2] HybridRAG: Integrating Knowledge Graphs and Vector Retrieval



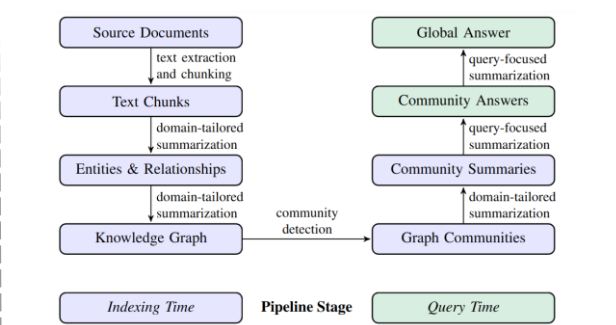
[3] CRAG: A Comprehensive Benchmark for RAG Evaluation



[4] Medical GraphRAG



[5] Benchmarking the Impact of Knowledge Graphs on LLM SQL QA



[6] GraphRAG: Global Sensemaking with Graph-Based Retrieval

Background

LLM only vs RAG vs GraphRAG

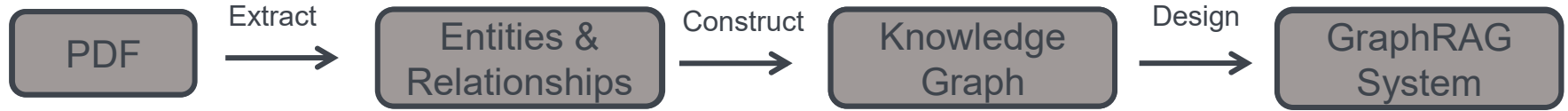
Criteria	LLM only	LLM + RAG	LLM + GraphRAG
Real-time Knowledge Access	Static, Pretrained	Dynamic , Unstructured	Dynamic, Structured
Domain Accuracy	Low (30%-60%) [3]	Improved (50%-80%) [1,2,3,6]	High (70%-90%) [2,4,5,6]
Explainable	Weak	Weak	Strong
System Complexity	Low	Moderate	High

System design

- System Overview
- Retrieval Enhancement

System design

System Overview



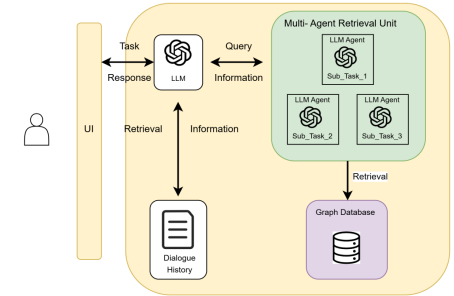
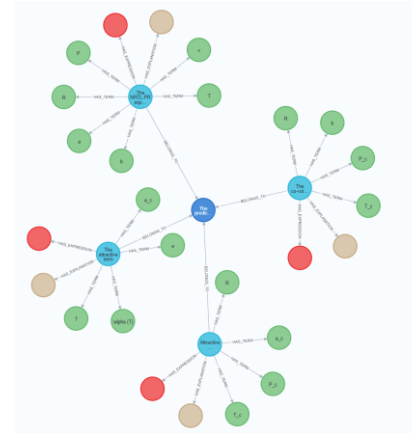
- Academic papers,
- Textbooks,
- Authoritative databases
- etc.

85 sets of structured data

- Formula,
- Expression,
- Explanation,
- Term,
- Source

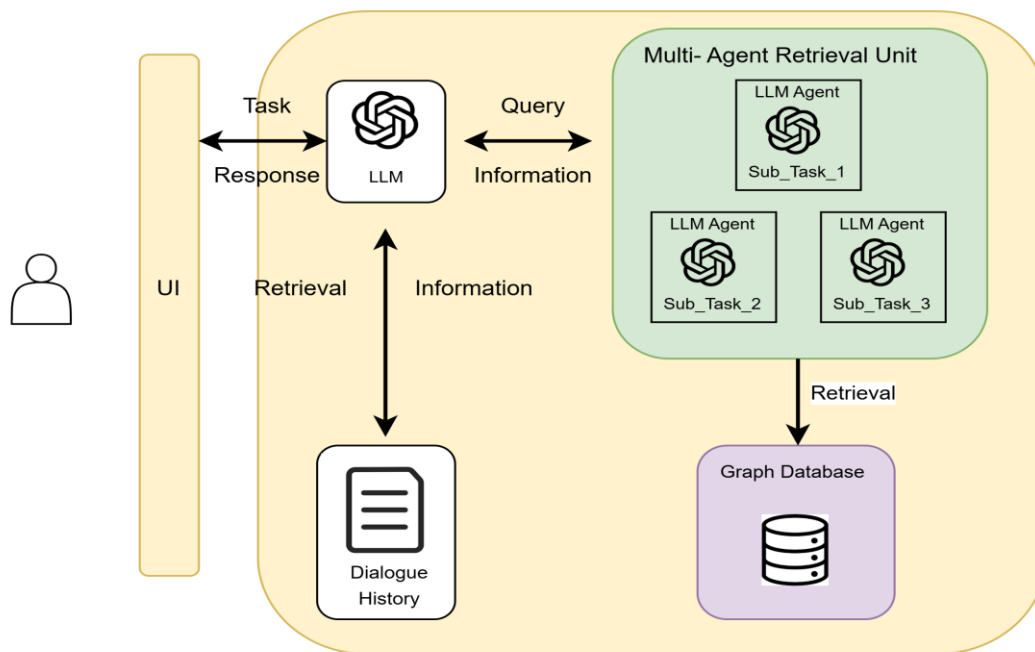
```

{
  "name": "A",
  "formula": "C12H18O6",
  "description": "A discretized population balance for nucleation growth and aggregation.pdf"
},
{
  "name": "B",
  "formula": "C12H18O6",
  "description": "A new formulation of the predictive NRTL-PR model in terms of kij mixing rules.pdf"
},
{
  "name": "C",
  "formula": "C12H18O6",
  "description": "A novel mixing rule model to predict the flowability of directly compressed pharma"
},
{
  "name": "D",
  "formula": "C12H18O6",
  "description": "Adjustable discretized population balance equations_Numerical simulation and para"
},
{
  "name": "E",
  "formula": "C12H18O6",
  "description": "Application of a System Model for Continuous Manufacturing of an Active Pharmaci"
},
{
  "name": "F",
  "formula": "C12H18O6",
  "description": "Brittle fracture in crystallization processes Part A. Attrition and abrasion of brittle so"
},
{
  "name": "G",
  "formula": "C12H18O6",
  "description": "Chord length distribution to particle size distribution.pdf"
},
{
  "name": "H",
  "formula": "C12H18O6",
  "description": "Kinetic modeling of aqueous and hydrothermal synthesis of barium titanate.pdf"
},
{
  "name": "I",
  "formula": "C12H18O6",
  "description": "Population balance modeling and optimization of an integrated batch crystallizer-w"
},
{
  "name": "J",
  "formula": "C12H18O6",
  "description": "Population balance modeling of growth and secondary nucleation by attrition and r"
},
{
  "name": "K",
  "formula": "C12H18O6",
  "description": "Secondary Nucleation by Interparticle Energies. II. Kinetics.pdf"
},
{
  "name": "L",
  "formula": "C12H18O6",
  "description": "Successful Formulation Window for the design of pharmaceutical tablets with requi"
}
  
```



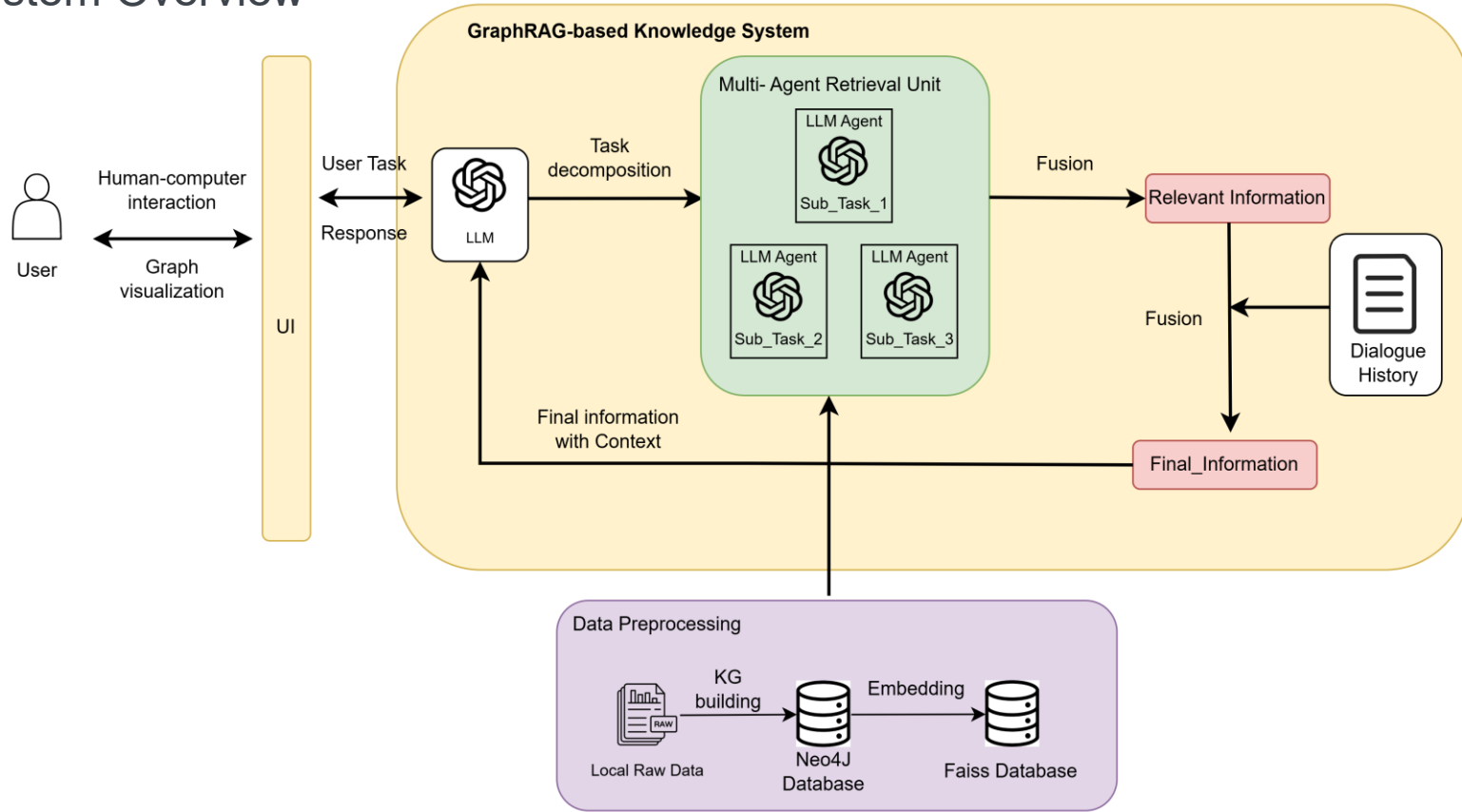
System design

System Overview



System design

System Overview



System design

- System Overview
- Retrieval Enhancement

GraphRAG

+ Chain of Thought

+ Semantic Alignment

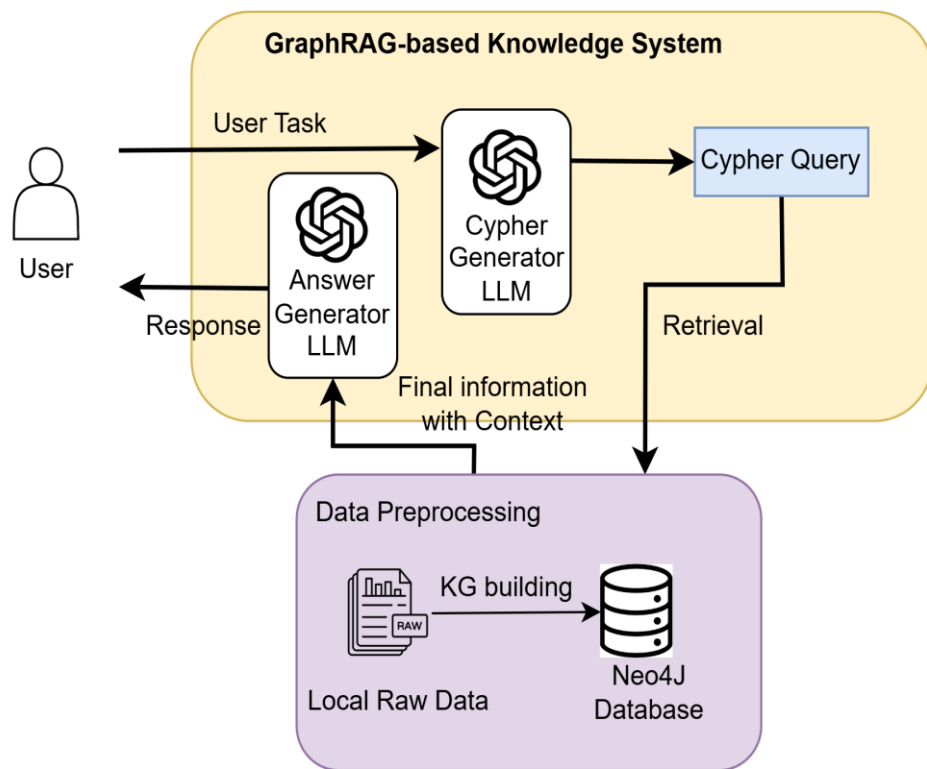
+ Verification

+ Task Decomposition

System design

Retrieval Enhancement

GraphRAG + CoT + Semantic Alignment + Verification + Task Decomposition



Task:

What is the **formula explanation** and **formula expression** of **Particle size balance model**?

Limitation

- Condition Omission
- Logical Confusion

System design

Retrieval Enhancement

CoT
Multi-condition Tasks Handling

GraphRAG + CoT + Semantic Alignment + Verification + Task Decomposition

What is CoT(Chain of Thought)?

Step-by-step reasoning process

Task:

What is the **formula explanation** and **formula expression** of **particle size balance model**?

CoT:

Step 1: Identify keywords from user task.

Step 2: Partial matching for each keyword.

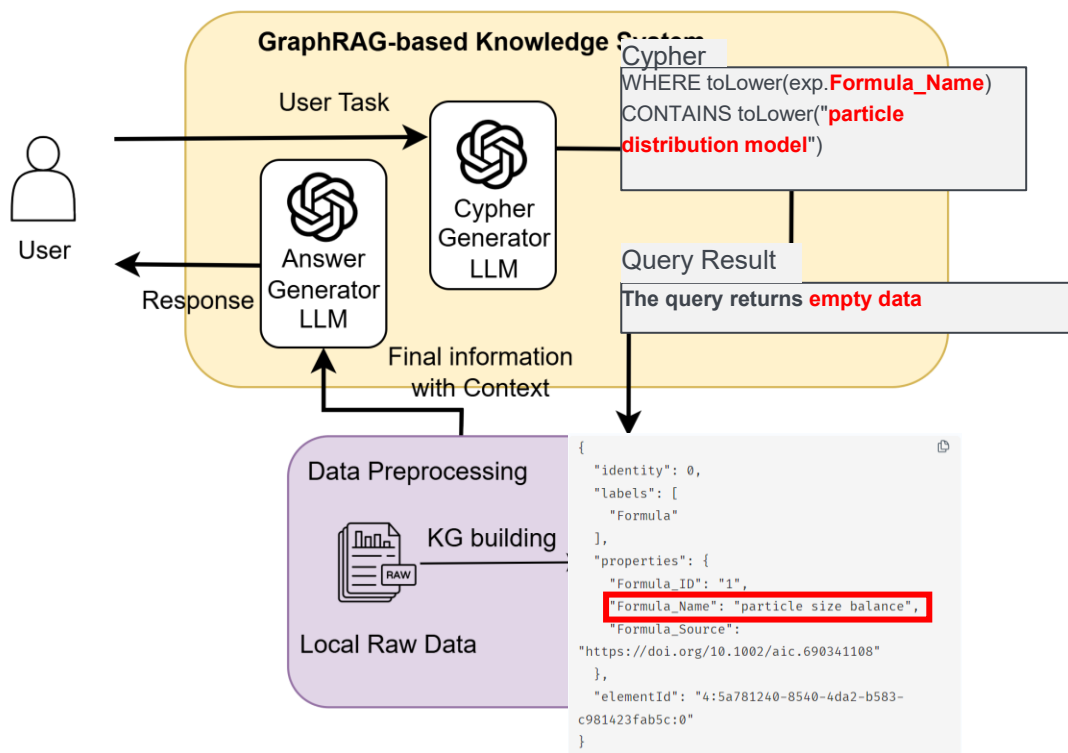
Step 3: Identify conditions and purposes.

Step 4: ...

System design

Retrieval Enhancement

GraphRAG + CoT + Semantic Alignment + Verification + Task Decomposition



CoT

Multi-condition Tasks Handling

Limitation

- Semantic Drift
- Vocabulary Inconsistency

Task:

What is the **explanation** of "**Particle distribution model**"?

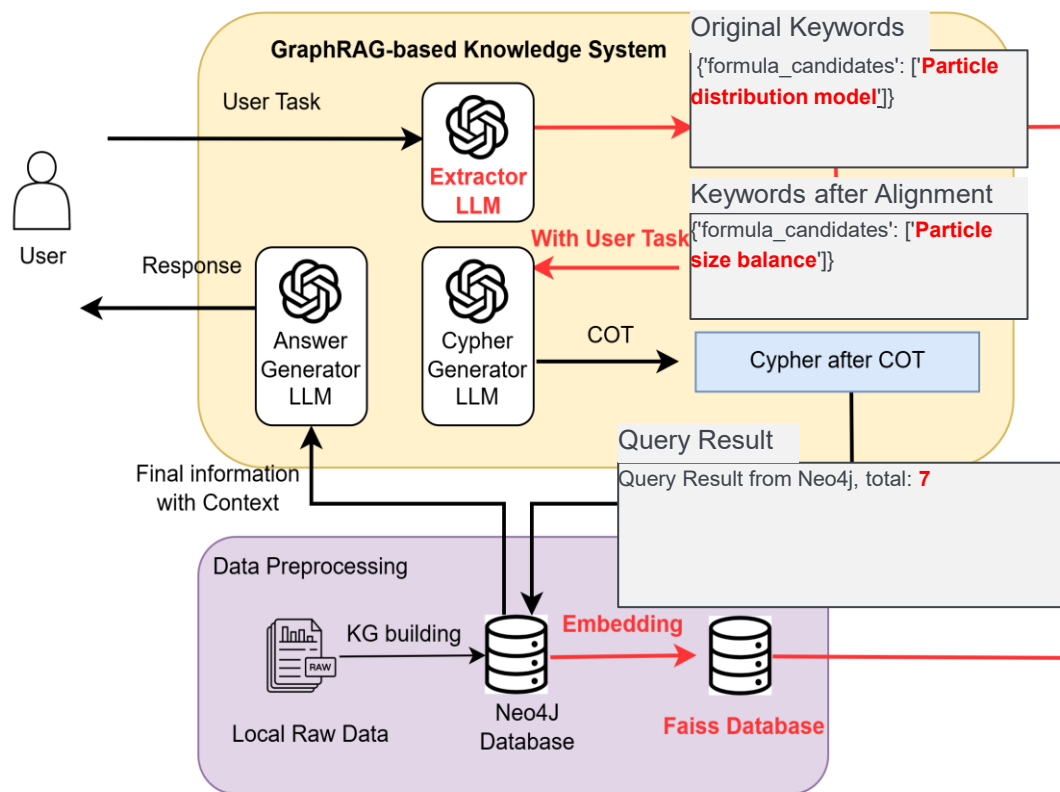
Answer:

The database does not contain relevant information.

System design

Retrieval Enhancement

GraphRAG + CoT + **Semantic Alignment** + Verification + Task Decomposition



Semantic Alignment
Semantic Drift & Vocabulary Inconsistency Mitigation

Task:

What is the **explanation** of "**Particle distribution model**"?

Answer:

The ****Particle distribution model****

Explanation of Formula:

This equation is commonly used to describe the dynamic behavior ...

Limitation

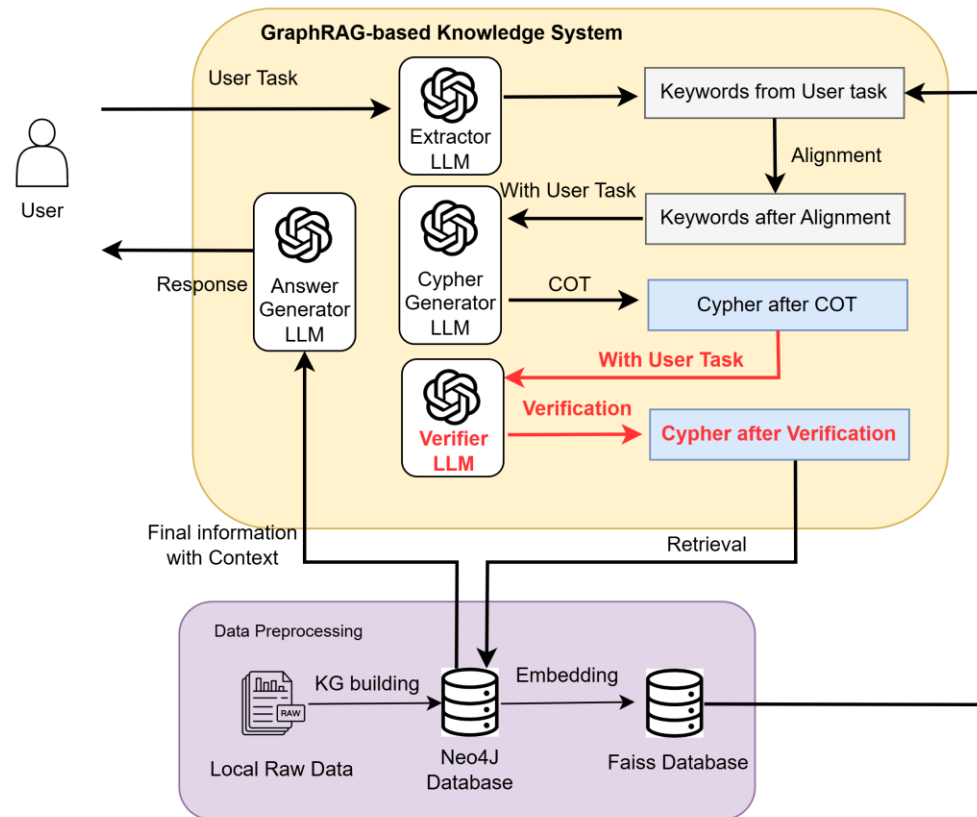
- Noisy output

System design

Retrieval Enhancement

GraphRAG + CoT + Semantic Alignment + **Verification** + Task Decomposition

Verification
Noisy output Mitigation



Check:

- **Syntax Correctness**
- **Logical Consistency**
- **Semantic Alignment**

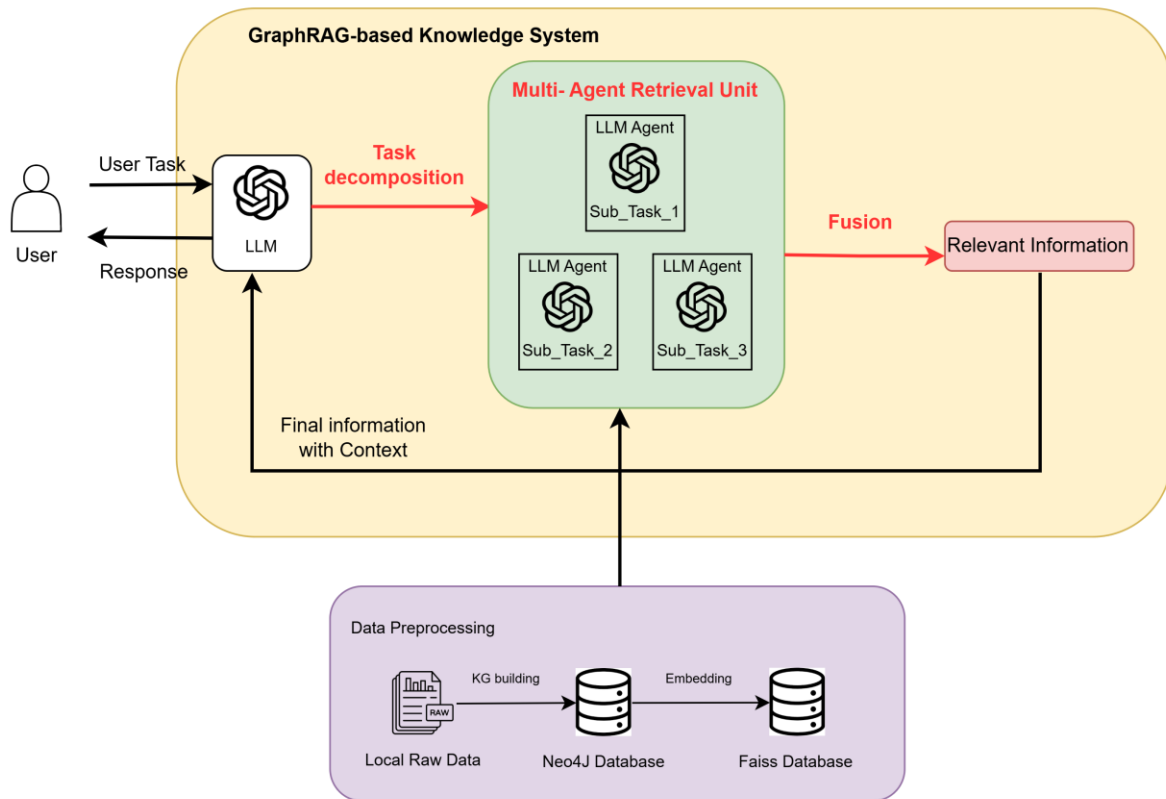
Limitation

- **Complex Tasks**

System design

Retrieval Enhancement

GraphRAG + CoT + Semantic Alignment + Verification + Task Decomposition



Task Decomposition Complex Tasks Handling

Task:

In the **particle size balance simulation process**, what factors determine the **birth item** and what factors are related to the **death item**?

Sub Task 1:

- "What factors determine the **birth item** in the **particle size balance simulation process**?"

Sub Task 2:

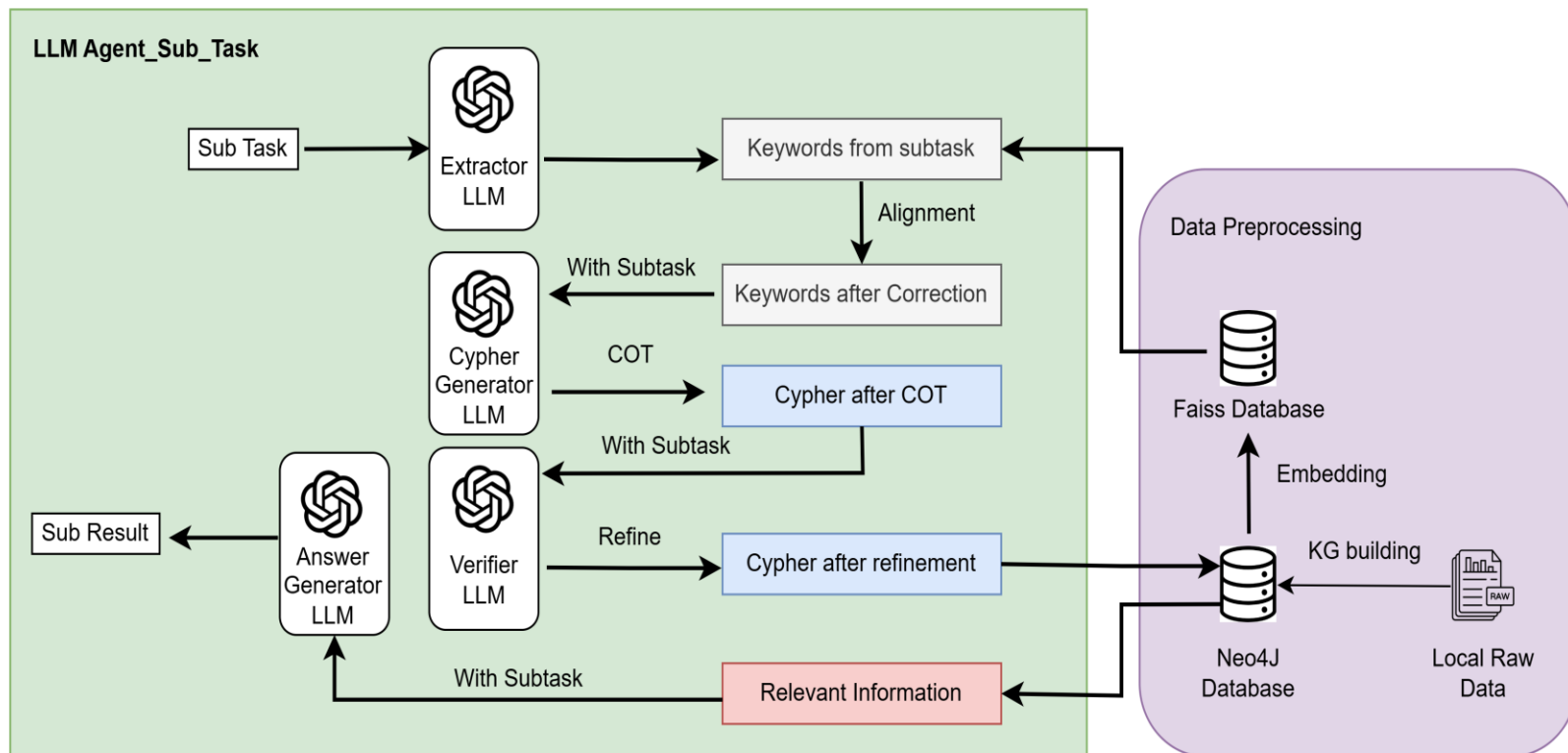
- What factors are related to the **death item** in the **particle size balance simulation process**?"

System design

Retrieval Enhancement

GraphRAG + CoT + Semantic Alignment + Verification + **Task Decomposition**

Task Decomposition
Complex Tasks Handling



Evaluation & Analysis

- Evaluation of information retrieval capability
- Evaluation of response generation quality

Evaluation & Analysis

Evaluation of information retrieval capability

Experiment Design

Step 1: Retrieval and Ground Truth generation

- 99 Questions
- For each method and question, compute:
 - Uniqueness and Verifiability

• **Hit Rate** =
$$\text{Hit Rate} = \frac{\text{Number of items correctly retrieved}}{\text{Total Number of items retrieved}}$$

Retrieval Strategy		Advantage
RAG	Vector-based semantic retrieval	Local Knowledge Utilization
Graph-based RAG	Basic GraphRAG	Relationship-driven Reasoning
GraphRAG + CoT	With Chain-of-Thought reasoning	Multi-condition Tasks Handling
GraphRAG + CoT + Semantic Alignment	+ Semantic Alignment	Semantic Drift Mitigation
GraphRAG + CoT + SA + Verification	+ Iterative query verification	Noisy output Mitigation
GraphRAG + CoT + SA + Verification + Task Decomposition	+ Task decomposition for multi-agent reasoning	Complex Tasks Handling

Evaluation & Analysis

Evaluation of information retrieval capability

Experiment Result & Analysis

Strategies Metrics	RAG	GraphRAG	GraphRAG +COT	GraphRAG +COT+SA	GraphRAG +COT+SA +Veri	GraphRAG +COT+SA +Veri+TD
Hit Rate	6.20%	25.50%	20.08%	21.90%	21.90%	14.95%
Recall	60.88%	58.06%	77.73%.	78.62%	82.69%	91.17%
Inference time	1.27s	2.79s	3.45s	5.67s	6.07s	8.20s

- Missing key knowledge is riskier than retrieving too much
- Strategy complexity ↑, Recall ↑, Precision↓, Inference time↑

Evaluation & Analysis

- Evaluation of information retrieval capability
- Evaluation of response generation quality

Evaluation & Analysis

Evaluation of response generation quality

Experiment Design

- LLM-based Judgment (Deepseek)
- Scoring Criteria

Metrics	Description
Accuracy Score(0-10)	Do the results match the user's expected answer?
Overfetching / Underfetching(0-10)	Does the result return too much irrelevant info or miss key data?
Readability(0-10)	Are the results understandable, structured, and informative?

Evaluation & Analysis

Evaluation of response generation quality

Experiment Result:

Strategies Metrics	RAG	GraphRAG	GraphRAG +COT	GraphRAG +COT+SA	GraphRAG +COT+SA +Veri	GraphRAG +COT+SA +Veri+TD
Accuracy Score	7.48	7.67	8.77	8.91	8.89	8.91
Overfetching / Underfetching	6.38	6.53	6.81	6.88	7.02	6.88
Readability	8.34	8.42	8.42	8.42	8.40	8.39

Conclusion & Outlook

Conclusion & Outlook

Overall

- Strong performance over RAG baseline
 - Graphrag: 91% recall vs RAG: 60%,
 - Improved precision and information completeness
- Limitations
 - 91% Recall, not 100% Coverage
 - Graph DB Construction Effort

Outlook

- Fusion with multimodal data
 - Text + Graph + Image , Cross-modal links
- Expansion into digital twin
 - Information Relationship Mapping
 - Digital twin environment



University of Stuttgart
Institut of Industrial Automation
and Software Engineering

Thank you!



Haonan Chen

e-mail st175447@stud.uni-stuttgart.de

phone +49 (0) 711 685-

fax +49 (0) 711 685-

University of Stuttgart
Institute of Automation and Software Systems
Pfaffenwaldring 47, 70550 Stuttgart

