# RBPA: Retrieval-Augmented-Generation based Personal Investment Assistant

**Xue Zeng***
PBC School of Finance
Tsinghua University
Beijing, China
zengx24@mails.tsinghua.edu.cn

**Yinuo Li***
Department of Construction Management
Tsinghua University
Beijing, China
liyinuo24@mails.tsinghua.edu.cn

## Abstract

RBPA employs Retrieval-Augmented Generation (RAG) to provide a sophisticated personal investment assistant. It addresses the limitations of Large Language Models (LLMs) in financial markets by integrating a knowledge base with real-time data retrieval. The system is designed to offer tailored investment advice by combining professional financial insights with individual investor data. Our approach includes a graph database for comprehensive document analysis and methods to enhance LLM's logical capabilities in finance. It aims to deliver personalized and informed investment strategies to users. The system also involves a comprehensive evaluation method and benchmark for investment decision, including both ROI and reasoning quality.

## 1 Introduction

### 1.1 Background

Large language models (**LLMs**) have made significant strides in recent years, with their capabilities being harnessed across a multitude of industries for various applications. These models, with their vast parameter counts, are designed to handle complex tasks and data, offering enhanced expressiveness and predictive performance.

Although LLMs can grasp basic world knowledge, they cannot be directly applied to financial markets in dynamic games. The influencing factors of the financial market are complex, from macro to micro and involve a wide range of aspects, and the analysis of the financial market needs to establish a professional knowledge base in the financial field, including basic professional knowledge, logical chain knowledge, and related network knowledge.

Retrieval-Augmented Generation (**RAG**) is a groundbreaking approach that marries the capabilities of large language models with the precision of information retrieval from reliable database. By leveraging a vast knowledge base, RAG enables the generation of highly accurate, relevant, and timely responses, making it an ideal technology for a personal investment assistant, under which circumstances lots of professional data and personal data are required.

### 1.2 Related Works

**RAG** Related works in the field of RAG have focused on enhancing the capabilities of large language models (LLMs) by incorporating external knowledge databases. Efforts like LangChain[1] and OpenAI's text-embedding-3 model[2] laid the foundation for vector-based retrieval systems. Works

*These authors contributed equally to this work.

Preprint. Under review.

Table 1: Data used in building graph database for RBPA.

| Data | Source | Data Volume | Daga | Data Describe |
|---|---|---|---|---|
| Finance News | CSMAR | 75k | 2018.01-2024.12 | We select news data directly related to stocks |
| Stock Price and Volume | Uqer | 7,498k | 2018.01-2024.12 | We use the stock's adjusted closing price data |
| Stock Limit Price | Uqer | 7,498k | 2018.01-2024.12 | To ensure the tradability of stocks, if the opening price of the next trading day hit the daily limit, we would exclude these transactions. |
| Analyst Report | Choice | 300 | 2018.01-2019.01 | |

such as the introduction of Palm by Chowdhery et al.[3] and the Mistral 7B model by Jiang et al.[4] also pushed the boundaries of scaling language models. Recent advancements in RAG systems, like Piperag by Jiang et al.[5], have emphasized the importance of fast retrieval-augmented generation. Meanwhile, studies on efficient vector search[6,7] and cache management strategies[8, 9] have aimed to optimize the performance of RAG workflows. Notably, the work of Kwon et al.[10] on PagedAttention has significantly improved memory management for LLM serving which is very important for RAG system for it usually takes very long inputs. Recently, Darren et al.[11] implemented graphRAG, which use LLM for database building and retrieve instead of traditional vector database and vector search.

**AI in investment** Kim and Nikolaev[12] exploit a artificial neural network combined with a large language model (BERT) to model the interpretation process in a way that allows us to directly capture the value of interactions between multimodal data. Citadel, a multinational hedge fund and financial services company with USD62 billion in assets under management, aims to include ChatGPT in its operations[13]. LLMs can discern intricate details from earnings reports to macroeconomic studies and process vast amounts of unstructured data, such as news articles or expert opinions, more efficiently than human analysts[14]. However, none of them can give personal investment suggestions for investors based on comprehensive information.

Therefore, to fill the gaps in the investment assistant field, we are going to present RBPA, a new system using RAG to help get professional data and personal data to generate better and more responsible suggestions for a investor.

## 2 Method

In order to help our system to gain the ability to generate responsible suggestions for a certain investor, we mainly did two jobs:

- build external database to get advanced knowledge of finance and the real-time information, as well as basic situation of the investor the system is now serving;
- help the LLM to get stronger logical ability on this specific field.

### 2.1 Graph database for RAG

Considering traditional vector database normally just split each document into several chunks, and saving the embedding of each chunk as the key of their corresponding text. This method of building vector database will easily lose many valuable information if they are far from each other. So, in order to keep long distance relationships and try to capture all useful information in each document, we will use graph database which contains different instances and relationships we collected from the documents. This process is designed to be done using LLM, with carefully designed prompt.

After the graph is built, we will use certain ways to merge similar information into groups, and use LLM to summarise each group in order to process multi-layer fast retrieve during inference.

We collected data from multi-sources, include finance news, stock price and volume, stock limit price and analyst report. Details about our data used can be found in table 1.

Based on the prepared data, we use GraphRAG Index-Engine[11] to build our graph database. We made some adjustments to the prompts used for the graph index, which you can find in Appendix for more details.

After 7 hours of indexing, we got our complete graph database, including a graph with 104073 nodes and 118097 edges and 8 clustered graph and summary for each community detected. We used Gephi to visualize the graph, using "Yifan Hu" layout and colored the nodes based on their entity type. The visualization result is showed in figure 1.

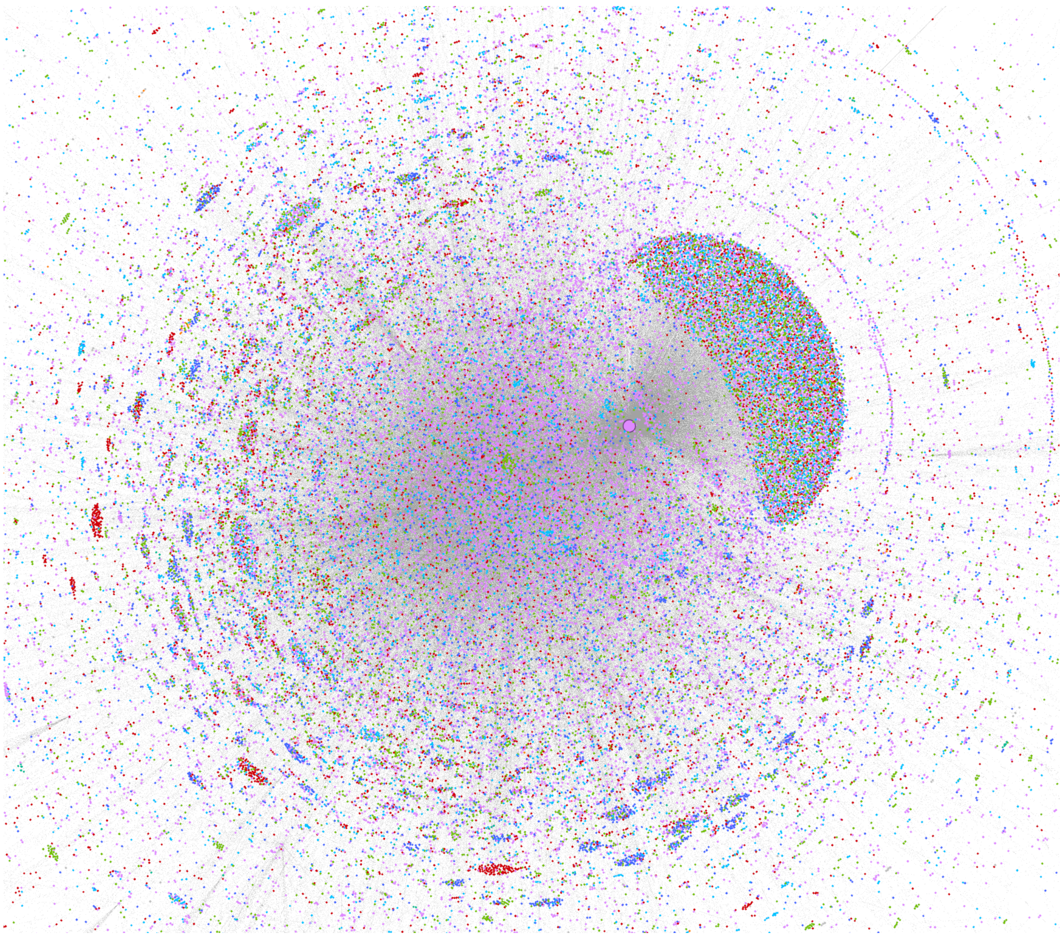


Figure 1: Visualization result using Gephi with "Yifan Hu" layout

### 2.2 Fine-tune base model

We use GLM-4-Flash as our base model. We fine-tuned the model through LoRA, using approximately 10,000 pieces of data as fine-tuning data, which included news data directly related to stocks and historical market data from August 1, 2023 to December 31, 2023. We have taken the following points into consideration.

#### 2.2.1 Superior Performance in Predictive Accuracy

GLM-4 demonstrates excellent performance in predicting outcomes accurately. It can handle complex relationships between variables and make precise predictions based on various types of input data. For example, in financial forecasting, it can analyze multiple economic indicators and historical data to provide more accurate predictions of stock prices or market trends compared to previous models.

### 2.2.2 Flexibility in Data Types and Structures

It is highly adaptable to different data types, including both continuous and categorical variables. Whether dealing with numerical data like sales figures or categorical data such as customer preferences, GLM-4 can effectively incorporate them into the model.

### 2.2.3 Fast Inference Speed

It achieves a stable speed of 72.14 tokens per second through techniques like adaptive weight quantization, parallel processing technology, batch processing strategies, and speculative sampling, which is outstanding among similar models.

## 2.3 Evaluation benchmark

In order to better evaluate the performance on investment suggestions, we implement an evaluation benchmark to comprehensively evaluate the return of investment and analyze quality of reasoning. The evaluation workflow is shown in figure 2

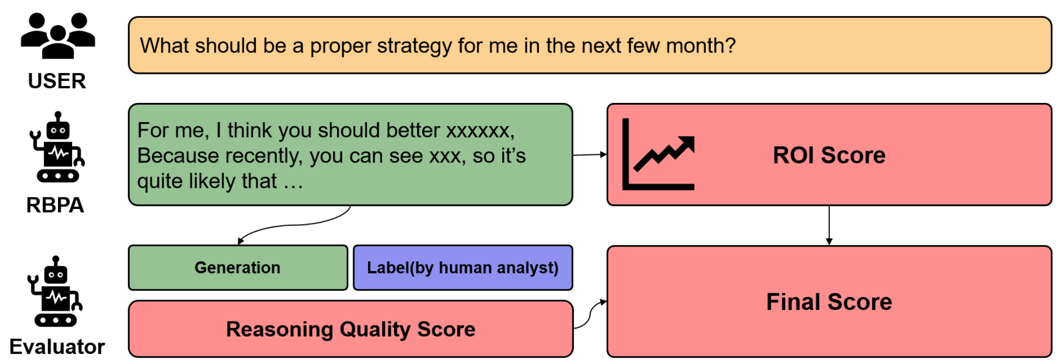



Figure 2: Workflow for evaluation benchmark, in order to eliminate the potential preference bias of LLM model for their own generation, we use a different LLM to evaluate the reasoning quality

### 2.3.1 ROI evaluation

We predict the increase or decrease range of a stock's price in the next five days by using the stock's past price data and the news on the current day. If the predicted value is greater than 1%, we will take a long position on the stock. If the predicted value is less than -1%, we will take a short position on the stock. And if the predicted value falls within the interval of [-1%, 1%], we will not take any action. In this way, we construct an investment portfolio. Based on the long-short portfolio return of the investment portfolio as well as the accuracy rate of the signals, and the recall rate of taking long and short positions, we use these as the evaluation indicators for the model's ability to predict stocks.

Considering that some news data were generated after the market closed, we used the opening price of the next trading day as the purchase price. To ensure the tradability of stocks, if the opening price of the next trading day hit the daily limit, we would exclude these transactions.

### 2.3.2 Reasoning quality evaluation

We use Qwen2-7B for the evaluation of reasoning quality. To stimulate the potential of this LLM, we further divide the quality into three aspects, including accuracy, depth and comprehensiveness. We then designed a prompt carefully to describe our task and the requirement for the output(json format), and add an example to better show what we expected.

We also emphasize our requirement for the evaluator to be strict, in case they give very high scores easily, which may cover the differences of our system with a general LLM. More details about this prompt can be found in Appendix.

Table 2: ROI evaluation results.

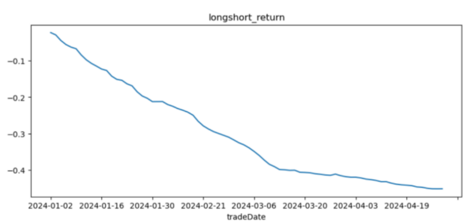
| | GLM-4-Flash | **Ours** |
|---|---|---|
| Accuracy rate | 0.29 | **0.42** |
| Precision-long | 0.12 | **0.13** |

Considering that the LLM may sometimes fail to follow the output json format, which will strongly influence our evaluation because the conversion from json string to python dictionary is very strict, we use a skip list during the evaluation to contain the index if the corresponding data is failed to be converted into python dictionary. After all data are processed, we go back to the data in the skip list and loop until it is empty.

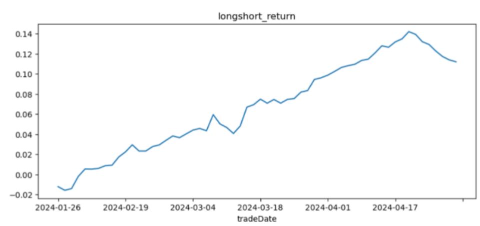
## 3 Experiments and Results

We use the base large model and the fine-tuned large model to conduct predictions and analyses on the out-of-sample data.The time interval of the out-of-sample data is from January to April 2024.

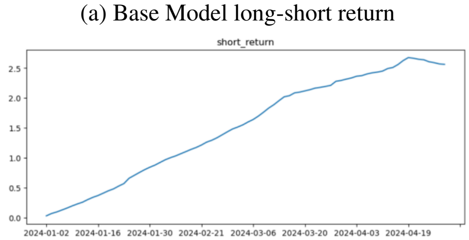
It can be seen from Table 2 that the accuracy rate of the base model is only 29%, while that of the fine-tuned model reaches 42%. According to Figure 3 (a) (b), the long-short portfolio from fine-tuned model achieves a positive return of 12%, and the investment portfolio obtained by the base model has a yield of -45%. The reason for the difference between the two lies in the long position investment portfolio. The fine-tuned large model is more accurate in predicting the long position.



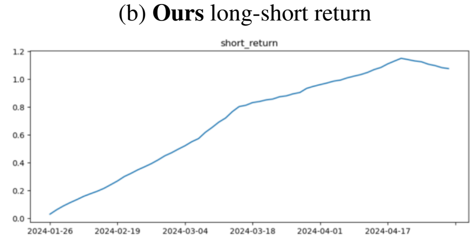

(a) Base Model long-short return



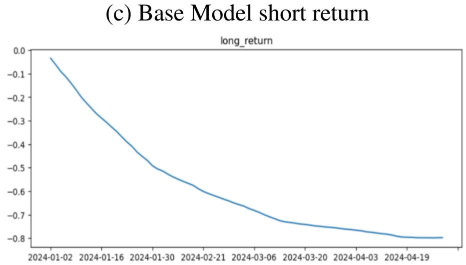

(b) **Ours** long-short return



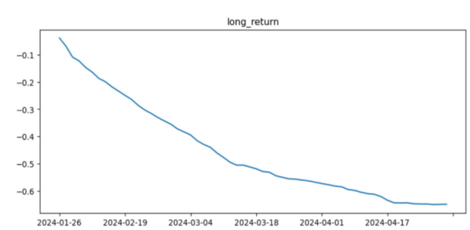

(c) Base Model short return




(d) **Ours** short return




(e) Base Model long return




(f) **Ours** long return

Figure 3: ROI Evaluation result visualization: The return of base model and our fine-tuned model under three different conditions, long-short, long and short.

Based on the output from the base model and our fine-tuned model, we collected the reasoning analyze and feed into the evaluator for reasoning quality evaluate, the result can be found in Figure 4.

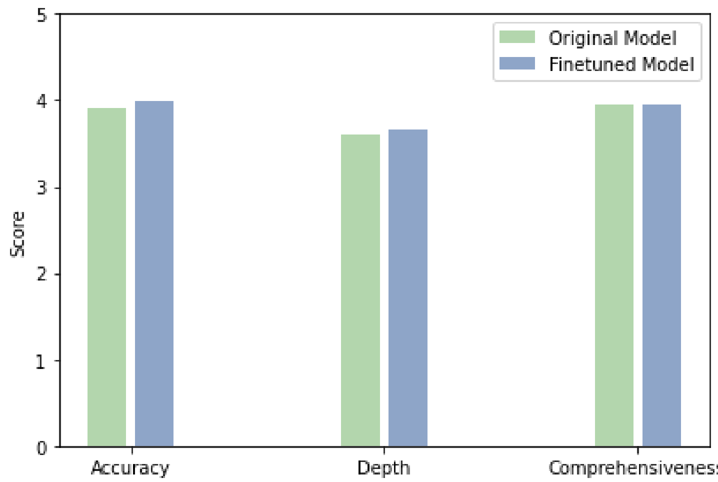



Figure 4: Reasoning Quality evaluate result, our model has better performance on "accuracy" and "depth" (2% and 1.5% respectively), and on the comprehensiveness metric, it is on par with the base model.

## 4 Conclusion

In conclusion, our Retrieval-Augmented-Generation based Personal Investment Assistant (RBPA) demonstrates a significant advancement in the field of personalized investment advice. By integrating Retrieval-Augmented Generation (RAG) with a comprehensive graph database, RBPA addresses the inherent limitations of Large Language Models (LLMs) in dynamic financial markets. Our system's ability to merge professional financial insights with individual investor data results in tailored investment strategies that are both personalized and informed.

The utilization of a graph database in RBPA has proven to be a pivotal enhancement, allowing for the preservation of long-distance relationships within documents and the capture of all valuable information. This approach has been instrumental in building a robust knowledge base that facilitates multi-layer fast retrieval during inference, which is crucial for real-time investment decisions.

Our fine-tuning approach, using GLM-4-Flash as the base model and fine-tuning through LoRA, has yielded promising results. The fine-tuned model exhibited a marked improvement in predictive accuracy, with an accuracy rate increase from 29% to 42%. The long-short portfolio returns also demonstrated the superiority of our fine-tuned model, achieving a positive return of 12%, compared to the base model's negative return of -45%.

Furthermore, our evaluation benchmark, which includes both ROI evaluation and reasoning quality assessment, provides a holistic view of the model's performance. The fine-tuned model outperformed the base model in terms of "accuracy" and "depth" of reasoning, solidifying its potential as a reliable personal investment assistant.

In summary, RBPA represents a substantial step forward in leveraging AI for personal investment assistance. It combines the strengths of RAG with a sophisticated graph database to deliver a system that not only meets the complex demands of financial analysis but also adapts to the unique needs of individual investors. Future work will focus on further enhancing the system's logical capabilities and expanding the knowledge base to cover an even broader range of financial data and scenarios, ensuring RBPA remains at the forefront of AI-assisted investment decision-making.

## References


[1] LangChain. (2024) https://python.langchain.com/docs/get_started/introduction

[2] OpenAI text-embedding-3 model. (2024) https://openai.com/blog/

[3] Aakanksha Chowdhery et al. (2022) Palm: Scaling language modeling with pathways. *arXiv preprint arXiv:2204.02311.*

[4] Albert Q Jiang et al. (2023) Mistral 7B. *arXiv preprint arXiv:2310.06825.*

[5] Wenqi Jiang et al.(2024) Piperag: Fast retrieval-augmented generation via algorithm-system co-design. *arXiv preprint arXiv:2403.05676.*

[6] Artem Babenko and Victor Lempitsky.(2014) The inverted multi-index. *IEEE transactions on pattern analysis and machine intelligence.*

[7] Yu A Malkov and Dmitry A Yashunin.(2018) Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. *IEEE transactions on pattern analysis and machine intelligence.*

[8] Tim Dettmers et al. (2022) Gpt3. int8 (): 8-bit matrix multiplication for transformers at scale. *Advances in Neural Information Processing Systems.*

[9] Zhuohan Li et al.(2020) Train big, then compress: Rethinking model size for efficient training and inference of transformers. in *International Conference on Machine Learning (ICML).*

[10] Woosuk Kwon et al.(2023) Efficient memory management for large language model serving with pagedattention. in *ACM SOSP.*

[11] Darren Edge et al.(2024) From Local to Global: A Graph RAG Approach to Query-Focused Summarization. *arXiv preprint arXiv:2404.16130.*

[12] Kim and Nikolaev.(2023) Context-Based Interpretation of Financial Information. *Chicago Booth Research Paper No. 23-08*

[13] Lin Tan, Huihang Wu and Xiaoyan Zhang.(2024) Large Language Models and Return Prediction in China. *SSRN 4712248, 2023*

[14] George Fatouros et al.(2024) Can Large Language Models Beat Wall Street? Unveiling the Potential of AI in Stock Selection. *arXiv preprint arXiv:2401.03737, 2024*

## Appendix

Here we show the prompts we used for graph database index and LLM evaluation.

### Graph Index prompts

#### claim extraction

-Target activity- You are an intelligent assistant that helps a human analyst to analyze claims against certain entities presented in a text document. \par -Goal- Given a text document that is potentially relevant to this activity, an entity specification, and a claim description, extract all entities that match the entity specification and all claims against those entities. \par -Steps- 1. Extract all named entities that match the predefined entity specification. Entity specification can either be a list of entity names or a list of entity types. \par 2. For each entity identified in step 1, extract all claims associated with the entity. Claims need to match the specified claim description, and the entity should be the subject of the claim. \par For each claim, extract the following information: - Subject: name of the entity that is subject of the claim, capitalized. The subject entity is one that committed the action described in the claim. Subject needs to be one of the named entities identified in step 1. - Object: name of the entity that is object of the claim, capitalized. The object entity is one that either reports/handles or is affected by the action described in the claim. If object entity is unknown, use **NONE**. \par - Claim Type: overall category of the claim, capitalized. Name it in a way that can be repeated across multiple text inputs, so that similar claims share the same claim type \par - Claim Status: **TRUE**, **FALSE**, or **SUSPECTED**. TRUE means the claim is confirmed, FALSE means the claim is found to be False, SUSPECTED means the claim is not verified. \par - Claim Description: Detailed description explaining the reasoning behind the claim, together with all the related evidence and references. \par - Claim Date: Period (start_date, end_date) when the claim was made. Both start_date and end_date should be in ISO-8601 format. If the claim was made on a single date rather than a date range, set the same date for both start_date and end_date. If date is unknown, return **NONE**. \par - Claim Source Text: List of **all** quotes from the original text that are relevant to the claim. \par Format each claim as (<subject_entity> {tuple_delimiter} <object_entity>{tuple_delimiter} <claim_type> {tuple_delimiter} <claim_status>{tuple_delimiter} <claim_start_date> {tuple_delimiter} <claim_end_date>{tuple_delimiter} <claim_description> {tuple_delimiter} <claim_source>) \par 3. Return output in English as a single list of all the claims identified in steps 1 and 2. Use **{record_delimiter}** as the list delimiter. \par 4. When finished, output {completion_delimiter} \par -Examples- Example 1: Entity specification: organization Claim description: red flags associated with an entity Text: According to an article on 2022/01/10, Company A was fined for bid rigging while participating in multiple public tenders published by Government Agency B. The company is owned by Person C who was suspected of engaging in corruption activities in 2015. Output: \par (COMPANY A{tuple_delimiter} GOVERNMENT AGENCY B{tuple_delimiter} ANTI-COMPETITIVE PRACTICES {tuple_delimiter} TRUE {tuple_delimiter} 2022-01-10T00:00:00 {tuple_delimiter} 2022-01-10T00:00:00 {tuple_delimiter} Company A was found to engage in anti-competitive practices because it was fined for bid rigging in multiple public tenders published by Government Agency B according to an article published on 2022/01/10 {tuple_delimiter} According to an article published on 2022/01/10, Company A was fined for bid rigging while participating in multiple public tenders published by Government Agency B.) {completion_delimiter} \par Example 2: Entity specification: Company A, Person C Claim description: red flags associated with an entity Text: According to an article on 2022/01/10, Company A was fined for bid rigging while participating in multiple public tenders published by Government Agency B. The company is owned by Person C who was suspected of engaging in corruption activities in 2015. Output: \par (COMPANY A {tuple_delimiter} GOVERNMENT AGENCY B {tuple_delimiter} ANTI-COMPETITIVE PRACTICES {tuple_delimiter} TRUE {tuple_delimiter} 2022-01-10T00:00:00 {tuple_delimiter} 2022-01-10T00:00:00 {tuple_delimiter} Company A was found to engage in anti-competitive practices because it was fined for bid rigging in multiple public tenders published by Government Agency B according to an article published on 2022/01/10{tuple_delimiter}According to an article published on 2022/01/10, Company A was fined for bid rigging while participating in multiple public tenders published by Government Agency B.) {record_delimiter} (PERSON C {tuple_delimiter} NONE{tuple_delimiter} CORRUPTION{tuple_delimiter} SUSPECTED {tuple_delimiter} 2015-01-01T00:00:00 {tuple_delimiter} 2015-12-30T00:00:00 {tuple_delimiter} Person C was suspected of engaging in corruption activities in 2015{tuple_delimiter}The company is owned by Person C who was suspected of engaging in corruption activities in 2015) {completion_delimiter} \par -Real Data- Use the following input for your answer. Entity specification: {entity_specs} Claim description: {claim_description} Text: {input_text} Output:

**entity extraction**

-Goal- Given a text document that is potentially relevant to this activity and a list of entity types, identify all entities of those types from the text and all relationships among the identified entities. \par -Steps- 1. Identify all entities. For each identified entity, extract the following information: - entity_name: Name of the entity, capitalized - entity_type: One of the following types: [{entity_types}] - entity_description: Comprehensive description of the entity's attributes and activities Format each entity as ("entity" {tuple_delimiter} <entity_name>{tuple_delimiter} <entity_type> {tuple_delimiter} <entity_description>) \par 2. From the entities identified in step 1, identify all pairs of (source_entity, target_entity) that are *clearly related* to each other. For each pair of related entities, extract the following information: - source_entity: name of the source entity, as identified in step 1 - target_entity: name of the target entity, as identified in step 1 - relationship_description: explanation as to why you think the source entity and the target entity are related to each other - relationship_strength: a numeric score indicating strength of the relationship between the source entity and target entity Format each relationship as ("relationship"{tuple_delimiter} <source_entity>{tuple_delimiter} <target_entity> {tuple_delimiter}<relationship_description> {tuple_delimiter} <relationship_strength>) \par 3. Return output in English as a single list of all the entities and relationships identified in steps 1 and 2. Use **{record_delimiter}** as the list delimiter. \par 4. When finished, output {completion_delimiter} \par ############################################# -Examples- ############################################# Example 1: Entity_types: ORGANIZATION,PERSON Text: The Verdantis's Central Institution is scheduled to meet on Monday and Thursday, with the institution planning to release its latest policy decision on Thursday at 1:30 p.m. PDT, followed by a press conference where Central Institution Chair Martin Smith will take questions. Investors expect the Market Strategy Committee to hold its benchmark interest rate steady in a range of 3.5############################################# Output: ("entity" {tuple_delimiter} CENTRAL INSTITUTION {tuple_delimiter} ORGANIZATION {tuple_delimiter} The Central Institution is the Federal Reserve of Verdantis, which is setting interest rates on Monday and Thursday) {record_delimiter} ("entity" {tuple_delimiter} MARTIN SMITH {tuple_delimiter} PERSON {tuple_delimiter} Martin Smith is the chair of the Central Institution) {record_delimiter} ("entity" {tuple_delimiter} MARKET STRATEGY COMMITTEE {tuple_delimiter} ORGANIZATION {tuple_delimiter} The Central Institution committee makes key decisions about interest rates and the growth of Verdantis's money supply) {record_delimiter} ("relationship" {tuple_delimiter} MARTIN SMITH {tuple_delimiter} CENTRAL INSTITUTION {tuple_delimiter} Martin Smith is the Chair of the Central Institution and will answer questions at a press conference {tuple_delimiter} 9) {completion_delimiter} \par ############################################# Example 2: Entity_types: ORGANIZATION Text: TechGlobal's (TG) stock skyrocketed in its opening day on the Global Exchange Thursday. But IPO experts warn that the semiconductor corporation's debut on the public markets isn't indicative of how other newly listed companies may perform. \par TechGlobal, a formerly public company, was taken private by Vision Holdings in 2014. The well-established chip designer says it powers 85############################################# Output: ("entity"{tuple_delimiter} TECHGLOBAL {tuple_delimiter} ORGANIZATION {tuple_delimiter} TechGlobal is a stock now listed on the Global Exchange which powers 85{record_delimiter} ("entity" {tuple_delimiter} VISION HOLDINGS {tuple_delimiter} ORGANIZATION {tuple_delimiter} Vision Holdings is a firm that previously owned TechGlobal) {record_delimiter} ("relationship" {tuple_delimiter} TECHGLOBAL {tuple_delimiter} VISION HOLDINGS {tuple_delimiter} Vision Holdings formerly owned TechGlobal from 2014 until present {tuple_delimiter} 5) {completion_delimiter} \par ############################################# Example 3: Entity_types: ORGANIZATION,GEO,PERSON Text: Five Aurelians jailed for 8 years in Firuzabad and widely regarded as hostages are on their way home to Aurelia. \par The swap orchestrated by Quintara was finalized when \$8bn of Firuzi funds were transferred to financial institutions in Krohaara, the capital of Quintara. \par The exchange initiated in Firuzabad's capital, Tiruzia, led to the four men and one woman, who are also Firuzi nationals, boarding a chartered flight to Krohaara. \par They were welcomed by senior Aurelian officials and are now on their way to Aurelia's capital, Cashion. \par The Aurelians include 39-year-old businessman Samuel Namara, who has been held in Tiruzia's Alhamia Prison, as well as journalist Durke Bataglani, 59, and environmentalist Meggie Tazbah, 53, who also holds Bratinas nationality. ############################################# Output: ("entity"{tuple_delimiter} FIRUZABAD {tuple_delimiter} GEO {tuple_delimiter} Firuzabad held Aurelians as hostages) {record_delimiter} ("entity"{tuple_delimiter} AURELIA{tuple_delimiter} GEO{tuple_delimiter} Country seeking to release hostages) {record_delimiter} ("entity"{tuple_delimiter} QUINTARA{tuple_delimiter} GEO{tuple_delimiter} Country that negotiated a swap of money in exchange for hostages) {record_delimiter} {record_delimiter} ("entity"{tuple_delimiter} TIRUZIA{tuple_delimiter} GEO{tuple_delimiter} Capital of Firuzabad where the Aurelians were being held) {record_delimiter} ("entity"{tuple_delimiter} KROHAARA{tuple_delimiter} GEO{tuple_delimiter} Capital city in Quintara) {record_delimiter} ("entity"{tuple_delimiter} CASHION{tuple_delimiter} GEO{tuple_delimiter} Capital city in Aurelia) {record_delimiter} ("entity"{tuple_delimiter} SAMUEL NAMARA {tuple_delimiter} PERSON {tuple_delimiter} Aurelian who spent time in Tiruzia's Alhamia Prison) {record_delimiter} ("entity" {tuple_delimiter} ALHAMIA PRISON {tuple_delimiter} GEO {tuple_delimiter} Prison in Tiruzia) {record_delimiter} ("entity"{tuple_delimiter} DURKE BATAGLANI {tuple_delimiter} PERSON {tuple_delimiter} Aurelian journalist who was held hostage) {record_delimiter} ("entity"{tuple_delimiter} MEGGIE TAZBAH {tuple_delimiter} PERSON {tuple_delimiter} Bratinas national and environmentalist who was held hostage) {record_delimiter} ("relationship"{tuple_delimiter} FIRUZABAD {tuple_delimiter} AURELIA {tuple_delimiter} Firuzabad negotiated a hostage exchange with Aurelia {tuple_delimiter} 2) {record_delimiter} ("relationship"{tuple_delimiter} QUINTARA {tuple_delimiter} AURELIA {tuple_delimiter} Quintara brokered the hostage exchange between Firuzabad and Aurelia{tuple_delimiter}2) {record_delimiter} ("relationship"{tuple_delimiter} QUINTARA {tuple_delimiter} FIRUZABAD {tuple_delimiter} Quintara brokered the hostage exchange between Firuzabad and Aurelia{tuple_delimiter}2) {record_delimiter} ("relationship" {tuple_delimiter} SAMUEL NAMARA{tuple_delimiter}ALHAMIA PRISON{tuple_delimiter}Samuel Namara was a prisoner at Alhamia prison{tuple_delimiter}8) {record_delimiter} ("relationship"{tuple_delimiter}SAMUEL NAMARA{tuple_delimiter}MEGGIE TAZBAH{tuple_delimiter}Samuel Namara and Meggie Tazbah were exchanged in the same hostage release{tuple_delimiter}2) {record_delimiter} ("relationship"{tuple_delimiter}SAMUEL NAMARA{tuple_delimiter}DURKE BATAGLANI{tuple_delimiter}Samuel Namara and Durke Bataglani were exchanged in the same hostage release{tuple_delimiter}2) {record_delimiter} ("relationship"{tuple_delimiter}MEGGIE TAZBAH{tuple_delimiter}DURKE BATAGLANI{tuple_delimiter}Meggie Tazbah and Durke Bataglani were exchanged in the same hostage release{tuple_delimiter}2) {record_delimiter} ("relationship"{tuple_delimiter}SAMUEL NAMARA{tuple_delimiter}FIRUZABAD{tuple_delimiter}Samuel Namara was a hostage in Firuzabad{tuple_delimiter}2) {record_delimiter} ("relationship"{tuple_delimiter}MEGGIE TAZBAH{tuple_delimiter}FIRUZABAD{tuple_delimiter}Meggie Tazbah was a hostage in Firuzabad{tuple_delimiter}2) {record_delimiter} ("relationship"{tuple_delimiter}DURKE BATAGLANI{tuple_delimiter}FIRUZABAD{tuple_delimiter}Durke Bataglani was a hostage in Firuzabad{tuple_delimiter}2) {completion_delimiter} \par -Real Data- \par Entity_types: {entity_types} Text: {input_text} \par Output:

**Evaluation prompt**

Hello, I need you to rate the quality of the following two analysis texts. The first paragraph is the analysis from a large model, and the second paragraph is the analysis from a human analyst. You will rate them based on the following three criteria and output the results in JSON format: \par Accuracy: The degree to which the analysis results match the known correct answers. Depth: Whether the analysis is in-depth and provides deep insights. Comprehensiveness: Whether the analysis is comprehensive and covers all relevant angles. Please give a rating from 1 to 5 for each criterion, where 1 indicates very poor and 5 indicates very good. Here are the two analysis texts: \par Large Model Output: {generation} \par Human Analyst: {label} \par To help you better understand the rating criteria and the JSON format output, here is a rating example \par Rating Example: """ { "accuracy": 4, "depth": 3, "comprehensiveness": 4, "reasons": { "accuracy": "The LLM's analysis is consistent with the human analyst's analysis in most cases, but deviates in some details.", "depth": "Although the LLM provides some in-depth analysis, it fails to delve into the root of the problem as deeply as the human analyst.", "comprehensiveness": "The LLM's analysis covers most key points, but is not as comprehensive as the human analyst in handling some edge cases." } } """