RERA: A MULTI-AGENT FRAMEWORK WITH TEXTUAL GRADIENT OPTIMIZATION FOR REAL ESTATE ENTERPRISE RISK ATTRIBUTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Real estate markets are highly susceptible to policy shifts and macroeconomic fluctuations, making accurate risk prediction and attribution essential for enterprise decision-making. Although large language models have shown considerable promise in processing multi-source data, integrating disparate information such as company financial reports, policy documents, and macroeconomic indicators into coherent risk transmission chains remains a persistent challenge. To bridge this gap, we introduce RERA, a multi-agent framework with textual gradient optimization for real estate enterprise risk attribution through large language models (LLMs). RERA employs an analyst-attribution agent architecture, in which analysts review policy, macroeconomic, and financial information to provide structured semantic interpretations and summarized analyses. The attribution agent consolidates insights from multiple analyst agents to identify external factors relevant to the company, establish causal chains, and generate comprehensive risk attribution reports. Its attribution optimization component evaluates generated reports against human expert reports for logical consistency, factual alignment, and quantitative reasonableness. These evaluation results are treated as textual gradients, and the attribution prompt is optimized accordingly to guide the agent toward more expert-aligned and evidence-grounded reports. Experiments on a multi-source dataset of listed real estate enterprises show that RERA outperforms general and finance LLMs on reference-based text metrics. Human evaluation further yields favorable assessments of its generated reports, while identifying a remaining gap relative to expert-written reports.

1 INTRODUCTION

The real estate market plays a leading role in the steady growth of GDP, and fluctuations in this market typically have an adverse impact on economic growth rates Nneji et al. (2013); Zhang et al. (2016). As a result, the financial sector devises policies tailored to macroeconomic conditions to regulate the real estate market Iacoviello (2005). This means that the risks faced by real estate enterprises are rarely isolated balance-sheet events. Instead, they are often influenced by policies, market demand, and macroeconomic expectations, which affect enterprises' financial performance through transmission mechanisms Gabauer & Gupta (2020). Although existing methods Zhu et al. (2023); Mena et al. (2024) can predict the level of a company's potential risk, decision makers need more than a binary warning about whether a risk may occur, but rather an understanding of how external factors give rise to financial risks.

Establishing such causal connections between macro-level conditions and firm-level outcomes is the central objective of risk attribution. By identifying risk drivers and clarifying the causal pathway from source to consequence, risk attribution supports risk assessment, early warning, and enterprise decision-making Li et al. (2025); Mwamba et al. (2025). In the real estate industry, the challenge of this task lies in integrating heterogeneous data from multiple sources, while ensuring that the conclusions drawn in the reports are supported by a complete causal chain and numeric consistency.

Despite the widespread adoption of large language models (LLMs) in finance, their application to real estate enterprise risk attribution and analysis remains limited due to the complexity of risk

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

drivers and the domain-specific nature of policy and macroeconomic texts Geerts et al. (2025); Cao et al. (2025). Moreover, using a single LLM for risk attribution makes data analysis, causal reasoning, and report writing implicit processes, which not only lack interpretability but also limit efficiency. Recent research suggests that organizing language agents into specialized roles can support complex tasks Chen et al. (2024b); Hong et al. (2023), offering a promising solution to the current limitations in risk attribution within the real estate industry.

In this paper, we introduce **RERA**, a multi-agent framework for **R**eal **E**state enterprise **R**isk **A**ttribution through LLMs. As shown in Figure 1, RERA follows an analyst-attribution architecture, complete with an attribution optimization component, to improve report quality. Drawing upon policy, macroeconomic, and corporate financial data provided by the task environment, analyst agents independently process these inputs to produce structured semantic interpretations and synthesized analyses. The attribution agent consolidates analysts' insights, identifies external factors relevant to the target company, and reasons through the causal chain, thereby attributing risks more accurately and generating a comprehensive report.

We further introduce an attribution optimization component based on textual gradient optimization Tang et al. (2025) to optimize the quality of generated reports. The component comprises an evaluation agent and a revision agent. The evaluation agent compares generated reports with human expert reports to assess logical consistency, factual accuracy, and quantitative reasonableness. This comparison identifies weaknesses in source selection, impact direction, causal-chain completion, and numerical grounding. The revision agent then expresses these weaknesses as textual gradients and converts them into bounded prompt updates for the attribution agent. This design keeps model parameters fixed, thereby preserving the analyst agents as stable evidence readers while enabling the attribution agent to gradually improve its selection of supported sources, connection of transmission channels, and grounding of numerical statements.

Our main contributions are summarized as follows:

- We develop **RERA**, an analyst-attribution agent framework for real estate enterprise risk attribution. RERA separates policy, macroeconomic, and financial analysis into dedicated analyst agents and integrates their outputs through an attribution agent to form traceable chains from source through channel to outcome.
- We design an attribution optimization component based on textual gradient optimization. It evaluates generated reports against human expert reports and converts identified weaknesses into textual gradients. These gradients are used to optimize the attribution agent prompt, thereby improving logical consistency, factual alignment, and quantitative reasonableness in generated attribution reports.
- We conduct comprehensive experiments on a constructed multi-source real estate enterprise risk attribution dataset. We compare RERA with general and finance LLMs using reference-based text metrics and assess its reports against expert-written reports through human evaluation. We further conduct ablation studies to examine the roles of textual gradient prompt optimization and staged attribution reasoning. Extensive experimental results demonstrate the effectiveness of RERA for real estate enterprise risk attribution.

2 Related Work

2.1 Real Estate Risk Attribution

Enterprises can use risk attribution to adopt appropriate risk management strategies and gain a competitive advantage Zarinjooei et al. (2025). For real estate enterprises, macro-level studies attribute fluctuations to housing demand, credit cycles, monetary policy, and uncertainty spillovers Shi et al. (2024); Chen & Wen (2017); Chen et al. (2024a). Policy-oriented studies further examine how taxation, macro-prudential regulation, coastal-market adjustment, and borrowing constraints reshape price dynamics and financing conditions Deng et al. (2022); McNamara et al. (2024). At the enterprise level, studies on policy and macro-prudential regulation connect external regulatory shocks to leverage, financing constraints, and systemic risk among real estate firms Ma & Xie (2025); Li et al. (2023); Chen et al. (2024c).

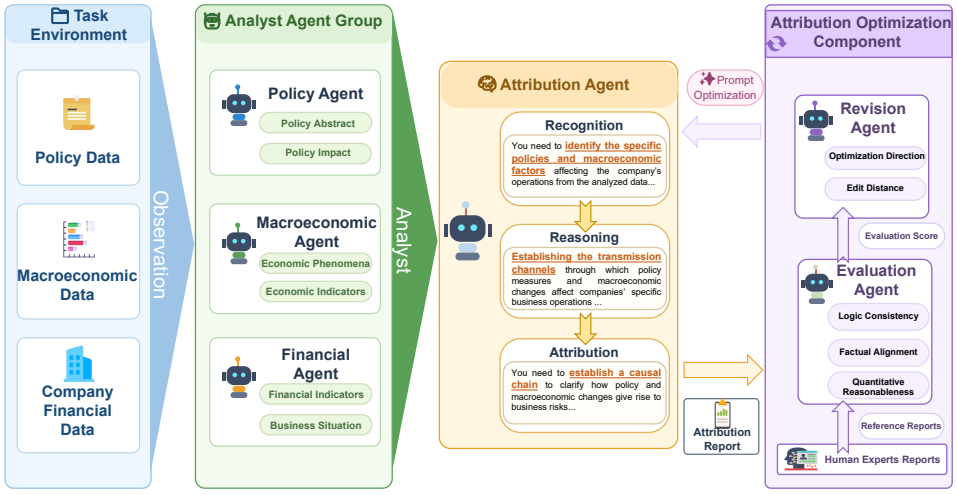


Figure 1: The overall framework of RERA.

Financial distress prediction and enterprise risk management research also provide warning indicators, symbolic classifiers, and risk-factor prioritization methods Zhu et al. (2023); Mena et al. (2024). However, crafting a comprehensive risk report requires a significant investment of manpower and time to read through a large volume of policy documents and to consolidate and analyze financial data. In this paper, we propose a multi-agent-based framework for generating comprehensive and accurate risk attribution reports, thereby effectively improving work efficiency.

2.2 LLM-BASED AGENTS

LLMs have been explored in finance for financial statement analysis, credit assessment, sentiment analysis, and investment screening Kim et al. (2024); Zaremba & Demir (2023); Peskoff & Stewart (2023); Delgadillo et al. (2024). Domain-specific models such as FinGPT Liu et al. (2023)and XuanYuan Zhang & Yang (2023) show that financial data construction and domain training can improve financial language understanding. Recent studies further extend LLMs to policy understanding, real estate appraisal, and multi-source financial risk prediction Geerts et al. (2025); Cao et al. (2025); Choi et al. (2026).

LLM-based agents provide a more structured alternative by decomposing complex tasks into role-based perception, reasoning, and action. For example, FinCon coordinates multiple financial agents with conceptual verbal reinforcement for financial decision-making Yu et al. (2024). StockAgent builds a simulated stock-trading environment to study LLM-based investor behavior Zhang et al. (2024), and market-simulation agents model interactions among heterogeneous financial actors Gao et al. (2024). Related surveys and public-policy agent systems show that role allocation can improve the transparency of complex economic reasoning Ding et al. (2024); Hao & Xie (2025).

For real estate enterprise risk attribution, however, the key challenge is to model the transmission chain from external environments to firm-level financial outcomes, while existing methods have not explicitly connected policy analysis with company financial analysis. To address this problem, RERA separates policy, macroeconomic, and financial evidence into dedicated analyst agents and uses an attribution agent to connect the evidence source, transmission channel, and risk outcome in a traceable attribution process.

2.3 PROMPT OPTIMIZATION AND VERBAL REINFORCEMENT

Reasoning-oriented prompting methods, such as chain-of-thought Wei et al. (2022); Wang et al. (2025), ReAct Yao et al. (2023b), and ART Paranjape et al. (2023), show that intermediate reasoning steps and tool use can make LLM agents' behavior more controllable in complex tasks. Another line of work directly optimizes prompts or agent behavior through natural-language feedback. Automatic prompt optimization and gradient-style prompt optimization treat model critiques as update

signals for prompts Pryzant et al. (2023); Tang et al. (2025), while Retroformer Yao et al. (2023a) and Reflexion Shinn et al. (2023) use retrospective or verbal feedback to improve language-agent decisions across trials. RERA adapts this idea to real estate risk attribution by using expert-report evaluation to identify shortcomings in the generated report. The attribution optimization component then converts these critiques into bounded prompt updates, making the Attribution Agent more consistent in selecting supported evidence, completing external-to-financial transmission explanations, and grounding financial claims.

3 METHOD

RERA is designed to generate enterprise risk attribution reports aiding enterprise decision-making based on the given analytical data. As shown in Figure 1, the framework consists of a task environment, an analyst agent group, an attribution agent, and an attribution optimization component. The analyst agents transform heterogeneous data into structured semantic insights. The attribution agent then consolidates these observations and generates the risk attribution reports. The attribution optimization component comprises the evaluation agent and the revision agent, using textual gradients derived from expert-report evaluation to update the attribution prompt during the training phase. All the prompts are listed in Appendix B.

3.1 TASK FORMULATION

For each real estate company c, the goal is to produce a risk attribution report analyzing the company's actual or potential financial risks and their causes, based on data from year t. The task input is a dataset comprised of external environmental information and internal company data:

$$X_{c,t} = \{x^p_{c,t}, x^m_{c,t}, x^f_{c,t}\}, \tag{1}$$

where $x^p_{c,t}$ denotes policy information, $x^m_{c,t}$ denotes macroeconomic information, and $x^f_{c,t}$ denotes company financial data. Policy and macroeconomic data provide candidate external causes. The financial data provides company-specific outcomes and numerical constraints. Given this dataset, the analyst agents group generates the corresponding summaries and analyses in textual format, i.e., $A_{c,t} = \{A^p_{c,t}, A^m_{c,t}, A^f_{c,t}\}$. The role of the attribution agent is to further integrate these insights, deduce risk transmission chains, and generate risk attribution reports:

$$y_{c,t} = G(A_{c,t}, p), \tag{2}$$

where $y_{c,t}$ denotes the generated report for company c in year t, and $G(A_{c,t}, p)$ represents the output from the attribution agent for input $A_{c,t}$ conditioned on the prompt p.

To analyze information accurately and improve the quality of attribution reports in multiple respects, the optimization component aims to optimize the attribution prompt p. Given the expert report $y^*_{c,t}$ for each firm-year case, the prompt-level objective can be written as:

$$p^* = \arg\max_{p \in \mathcal{P}_r} \mathbb{E}\left[F\left(G(A_{c,t}, p), y^*_{c,t}\right)\right], \tag{3}$$

where $\mathcal{P}_r$ represents the set of candidate prompts generated by the optimization component, and $F(\cdot)$ denotes the task performance based on the designed measurement and human expert $y^*_{c,t}$. By converting expert-report evaluation into textual gradients, RERA updates the attribution agent's capability to optimize later attribution reports.

3.2 ANALYST AGENT GROUP

In RERA, analyst agents extract key insights from external environmental information and company financial data, with each analysis focusing on the financial risks an enterprise potentially faces or is currently facing. To ensure high-quality reasoning by reducing task load and sharpening focus, each agent processes information from a single source and provides pre-specified outputs based on prompts. The role of the analyst agent group is to turn heterogeneous inputs into concise, typed observations that the attribution agent can understand. Each observation records the source, summarized content, expected impact direction, and supporting evidence.

Policy Agent. This agent reads real estate-related policy documents and extracts key information,

namely the policy abstract and the policy impact. The policy abstract outlines the source of the policy and its key provisions, and lists keywords such as "purchase restrictions", "reduced deposit requirements" and "financing restrictions". The policy impact describes the areas of a company's business that may be affected by policy, such as sales, cash flow, and tax burden. The agent will identify the affected areas and the expected direction of these impacts, providing a policy context for analyzing the enterprise's risks.

Macroeconomic Agent. To analyze the potential impact of macroeconomic conditions on the enterprise, the macroeconomic agent interprets macroeconomic data to distill existing macroeconomic phenomena and changes in economic indicators. Macroeconomic phenomena describe the current market conditions reflected in the data, such as contracting demand, loose liquidity, and weak price expectations. Economic indicators, meanwhile, provide the specific numerical evidence underpinning these phenomena, including the interrelationships between demand, total lending, property prices, and land market variables.

Financial Agent. By extracting enterprise financial indicators and summarising corresponding business performance, the financial analyst agent transforms financial data into enterprise-level evidence for attribution analysis. Financial indicators include leverage ratios, short-term debt pressures, operating cash flow and profit margins, and so on. The business situation describes the operational conditions reflected by these indicators, such as a lagging sales recovery, liquidity constraints, refinancing pressures, and the burden of project delivery. This agent monitors changes in a company's key indicators across different quarters within a year to analyze whether financial risks such as a contraction in scale or a decline in solvency are present. It links any anomalous financial changes to the company's current stage of development and ensures the accuracy of the figures in its analysis, thereby providing high-quality enterprise-level evidence for risk attribution.

3.3 Attribution Agent

The attribution agent receives the analyst observations $A^p_{c,t}$, $A^m_{c,t}$, and $A^f_{c,t}$ and generates the final report $y_{c,t}$ under the attribution prompt p. For LLM-based agents, dividing a task into multiple stages can enhance the agent's comprehension and the quality of task completion Plaat et al. (2024); Han et al. (2025). In this paper, to improve the accuracy and completeness of attribution reports, we divide the attribution agent's task into three stages: recognition, reasoning, and attribution.

Recognition. First, the attribution agent selects the policy or macroeconomic phenomena that are relevant to the target enterprise. Each selected item is supported by $A^p_{c,t}$ or $A^m_{c,t}$ and preserves the original source keyword or indicator description. This mechanism helps prevent unsupported expansion beyond the task environment and creates an auditable list of candidate causes.

Reasoning. The role of the second stage is to build the impact path based on each recognized factor. A valid path must link the external source to the financial indicator and then specify the changes that occurred in the business situation. The agent analyses changes in the enterprise's financial condition based on the $A^f_{c,t}$ and, taking into account information on the external environment, explains how a policy measure or macroeconomic phenomenon affects a concrete business area such as sales, financing, or cash collection.

Attribution. Lastly, the agent further extends the impact pathway to the enterprise's overall financial risk and analyzes the possible outcomes of the risk. The agent specifies how the changes in the business situation contribute to a concrete financial or operational risk, such as liquidity pressure, profitability decline, inventory accumulation, refinancing stress, or delivery burden. Thus, the complete chain of transmission from the external environment to enterprise risk is established. Following these three stages of analysis, the agent outputs the results in a structured format as a risk attribution report.

The risk attribution report generated by RERA covers five fields: data type, specific source, content summary, impact path, and risk impact. The data type field distinguishes whether the cause comes from policy or macroeconomic phenomena. The specific source field records the concrete policy keyword, macroeconomic indicator, or market signal used as the attribution basis. The content summary field states the external condition in a compact form. The impact path field explains the risk transmission mechanism from the external factor to the firm, for example from demand contraction to slower sales conversion and delayed cash collection. The risk impact field names the final firm-level consequence and links it to financial indicator evidence, such as operating cash flow pressure, profitability compression, or inventory turnover risk.

3.4 Attribution Optimization Component

We have innovatively designed an attribution optimization component to help identify and analyze shortcomings in the attribution process, thereby enhancing the capabilities of three stages of the attribution agent. The component is implemented by the evaluation agent and the revision agent. The evaluation agent compares a generated report with the expert report and expresses the gap as evaluation scores. The revision agent regards evaluation scores as the gradients and converts these gradients into bounded prompt updates.

3.4.1 Evaluation Agent

The optimization signal should be designed around overall attribution quality rather than the degree of word matching. Enterprise risk management studies also emphasize financial indicators, environmental indicators, and risk reasoning as core evaluation criteria Zarinjooei et al. (2025). Based on relevant academic research Jin et al. (2025), we design three task-specific dimensions to evaluate each generated attribution report before prompt revision. *Logic consistency* (LC) is used to assess whether the report closes the causal risk chain. *Factual alignment* (FA) denotes whether the selected policy or macroeconomic source and its impact direction match the expert report and the input evidence. *Quantitative reasonableness* (QR) checks whether numerical statements are directly supported by financial evidence and remain economically plausible. The score ranges for all three dimensions are set to 0–100.

Given a generated report $\hat{y}$ and the expert report y^*, the evaluation agent returns the three dimensions scores and the textual gradient:

$$(\mathbf{s}, g) = E(\hat{y}, y^*), \tag{4}$$

where $E(\cdot)$ denotes the evaluation process of the evaluation agent, and $\mathbf{s} = (s_{\mathrm{LC}}, s_{\mathrm{FA}}, s_{\mathrm{QR}})$ denotes the score vector corresponding to LC, FA, and QR. The textual gradient g summarizes the main error pattern, such as a missing transmission channel, a reversed impact direction, or an incorrect numerical analysis. $(\mathbf{s}, g)$ is then dispatched to the revision agent, while g indicates the direction in which the currently attributed prompt should be optimized, and $\mathbf{s}$ determines the extent to which the revision agent revises the current prompt.

3.4.2 Revision Agent

The revision agent generates candidate prompts based on $(\mathbf{s}, g)$ to update the attribution prompt. Specifically, in each iteration, the revision agent receives the scores and textual gradients of several reports and generates a new candidate prompt set $\mathcal{P}_r$ based on the optimization trajectory of the attribution prompt.

To control edit distance between updates at consecutive iterations, we apply a word-level edit budget with cosine decay Tang et al. (2025); Gotmare et al. (2019) in RERA:

$$\rho_t = \rho_{\min} + \frac{1}{2}(\rho_{\max} - \rho_{\min}) \left(1 + \cos \frac{\pi t}{T}\right), \tag{5}$$

where T is the maximum number of optimization steps, ρ_t denotes the current learning rate. The actual revision distance is adjusted by the mean evaluation score, so lower-scoring prompts receive stronger but still budget-bounded edits and higher-scoring prompts receive smaller refinements. Candidate prompts that satisfy this adjusted budget are used to generate new reports. The evaluation agent then outputs the new reports' score to evaluate the effectiveness of each candidate prompt. In the next iteration, RERA utilizes the prompt that yields the largest improvement in the mean evaluation score. If there is no improvement, the current prompt is retained.

4 Experiments

The experiments address four research questions: **RQ1**: How does RERA compare with general and finance LLMs in generating risk attribution reports? **RQ2**: Does evaluator-guided prompt revision improve report quality? **RQ3**: How do the recognition and reasoning stages affect report quality? **RQ4**: How does the evaluation agent's base LLM affect RERA's performance?

Method	BERTScore	SMART-1	SMART-L	ROUGE-1	ROUGE-L
General LLMs					
Gemini-2.5-pro	0.8433	0.3018	0.1650	0.4096	0.1779
GPT-4o	0.8853	0.2650	0.1553	0.4317	0.2007
GPT-4o-mini	0.8767	0.2809	0.1816	0.4120	0.2457
Qwen-14B-Instruct	0.8723	0.2105	0.1607	0.3451	0.1754
Llama3.1-8B-Instruct	0.8758	0.1892	0.1217	0.3048	0.1621
GLM4-9B-chat	0.8618	0.1708	0.1108	0.2025	0.1046
Finance LLMs					
FinGPT-mt-llama3-8B-lora	0.8753	0.2974	0.1882	0.3957	0.2185
XuanYuan-13B-chat	0.8601	0.3293	0.1995	0.4093	0.2777
FinGPT-ChatGLM2-6B	0.7820	0.2321	0.1309	0.2014	0.1545
TongYi-Finance-14B-chat	0.8977	0.3315	0.2341	0.4212	0.2954
RERA (Ours)	**0.9226**	**0.3780**	**0.2823**	**0.4702**	**0.3644**

Table 1: Comparison of RERA and LLM baselines on reference-based text metrics.

4.1 Dataset

We construct a multi-source dataset of 528 annual observations from 88 listed real estate companies over 2018–2023. Financial statements and annual report analyses are collected from company websites and Akshare[1]. Each firm-year sample combines policy, macroeconomic, and company financial information, with a reference risk attribution report prepared by a senior financial analyst. Appendices A.1 and A.2 detail data collection and distribution.

4.2 Experiment Setting

Baselines. We compare RERA with general and finance LLMs. General LLM baselines include Gemini-2.5-pro Comanici et al. (2025), GPT-4o and GPT-4o-mini OpenAI (2024), Qwen-14B-Instruct Bai et al. (2023), Llama3.1-8B-Instruct Llama Team, AI at Meta (2024), and GLM4-9B-chat GLM Team et al. (2024). Finance LLM baselines include FinGPT-mt-llama3-8B-lora[2], XuanYuan-13B-chat Zhang & Yang (2023), FinGPT-ChatGLM2-6B[3], and TongYi-Finance-14B-chat[4]. These baselines distinguish general language capabilities from financial specialization.
Implementation Details. Open-source models are deployed locally using Python libraries, while closed-source models are accessed through their official APIs. Unless otherwise specified, all agents in RERA use GPT-4o. The temperature is set to 0.3 for all LLMs. We use 70% of the dataset for training and optimize RERA for 3 epochs with a batch size of 32. The cosine decay schedule uses $\rho_{\max} = 0.30$ and $\rho_{\min} = 0.0$, and the revision agent generates 5 candidate prompts at each optimization step.

4.3 Evaluation Metrics

We evaluate reference alignment and domain-specific report quality.
Basic Metrics. Following the distinction between text-generation and domain-oriented evaluation Jin et al. (2025), we report BERTScore, SMART-1, SMART-L, ROUGE-1, and ROUGE-L Zhang et al. (2020); Lin (2004); Amplayo et al. (2022). BERTScore measures contextual semantic similarity to expert reports. SMART-1 and SMART-L assess sentence-level correspondence, while ROUGE-1 and ROUGE-L measure unigram and longest-common-subsequence overlap, respectively. Reference alignment does not independently establish factual accuracy or causal correctness.
Human Metrics. Human evaluation covers four dimensions adapted from enterprise risk management criteria Jin et al. (2025); Zarinjooei et al. (2025). *Financial Metrics* assesses numerical accuracy and the depth of financial analysis. *External Environment* assesses the relevance and completeness of contextual evidence, including policies, macroeconomic conditions, and company financial

[1] https://akshare.akfamily.xyz
[2] https://huggingface.co/FinGPT/fingpt-mt_llama3-8b_lora
[3] https://huggingface.co/FinGPT/fingpt-mt_chatglm2-6b_lora
[4] https://www.modelscope.cn/models/TongyiFinance/Tongyi-Finance-14B-Chat

characteristics. *Risk* assesses the plausibility and completeness of risk transmission chains. *Writing* assesses coherence, readability, and logical consistency. Each dimension is scored from 0 to 5.

4.4 Main Results

4.4.1 Main Results of Basic Metrics

Table 1 addresses RQ1. RERA achieves the highest scores on all five metrics. Compared with TongYi-Finance-14B-chat, the strongest baseline on BERTScore, SMART-L, and ROUGE-L, RERA improves these scores by 2.8%, 20.6%, and 23.4%, respectively, relative to the corresponding baseline score. The improvements span contextual semantics, sentence-level correspondence, and lexical alignment with expert references.

These results support the complementary roles of RERA's agents: dedicated analysts organize policy, macroeconomic, and financial evidence, while the attribution agent connects external factors to transmission channels and firm-level outcomes. RERA also outperforms its GPT-4o backbone used alone, suggesting that this structured workflow and prompt optimization contribute beyond base-model capabilities. Financial specialization by itself yields mixed results: TongYi-Finance-14B-chat leads most baseline metrics, whereas GPT-4o leads ROUGE-1. The ablation study further examines the contribution of each framework component.

4.4.2 Human Evaluation

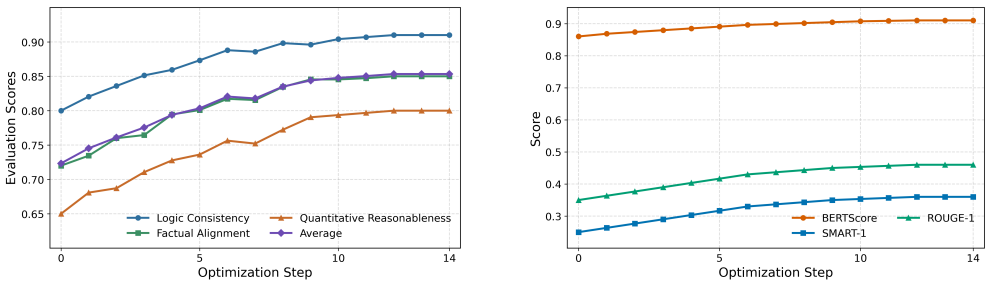
Three senior analysts assess 30 expert reports and their RERA-generated counterparts. Table 2 reports an average rating of 4.30 out of 5 for RERA and 4.60 for expert reports. RERA exceeds 4.2 in every dimension, indicating favorable assessments within this sample; experts score higher throughout.

Report Source	Financial	External	Risk	Writing	Average
Human expert written	4.62	4.58	4.66	4.55	4.60
RERA generated	4.28	4.31	4.26	4.34	4.30
Kappa score	0.85	0.86	0.89	0.84	0.86

Table 2: Human evaluation of RERA-generated and expert-written risk attribution reports.

Fleiss' kappa ranges from 0.84 to 0.89 (mean 0.86), indicating high inter-rater agreement Landis & Koch (1977). Writing is RERA's highest-rated dimension (4.34), whereas Risk is its lowest (4.26); expert reports show the opposite pattern for these two dimensions. This suggests that readable presentation is a relative strength, while the plausibility and completeness of risk transmission chains warrant further improvement. Report examples appear in Appendix C.

4.4.3 Optimization Process Analysis



(a) Evaluation scores

(b) Basic Metrics

Figure 2: Trends in evaluation scores and basic metrics during attribution prompt optimization.

To assess whether successive textual-gradient prompt updates improve report quality (RQ2), we track evaluator scores and reference-based metrics (Figure 2), examining changes across evaluation dimensions and whether performance stabilizes. Evaluator scores rise with minor fluctuations before stabilizing; logical consistency remains strongest, while quantitative reasonableness improves

but still lags. BERTScore, SMART-1, and ROUGE-1 also increase and plateau, indicating closer alignment with expert reports.

Together, these trends suggest that feedback on missing transmission channels, unsupported impact directions, and weak numerical explanations helps refine attribution prompts. The persistent weakness in quantitative reasonableness points to a need for stronger numerical grounding. Since evaluator scores guide prompt selection, their improvement should be considered alongside reference-based metrics and ablations. Appendix D further illustrates the optimization process.

4.5 Ablation Study

To address RQ2 and RQ3, we compare the full framework with three variants that separately remove attribution optimization, the recognition stage, or the reasoning stage.

Table 3 shows that all three removals reduce every reported metric. Relative to full RERA, removing attribution optimization reduces SMART-L by 14.3%, compared with 7.0% for recognition and 3.3% for reasoning. Removing optimization also causes the largest decrease on BERTScore and ROUGE-L, supporting the value of evaluator-guided prompt revision for reference alignment.

Table 3: Ablation results.

Variant	BERTScore	SMART-L	ROUGE-L
RERA	**0.9226**	**0.2823**	**0.3644**
w/o optimization	0.8583	0.2418	0.3127
w/o recognition	0.8797	0.2625	0.3351
w/o reasoning	0.8984	0.2731	0.3547

The recognition ablation suggests that selecting relevant external evidence is important before constructing transmission paths. Reasoning provides an additional benefit by connecting these factors to firm-level financial conditions. The recognition phase establishes the evidence basis, and the reasoning phase organizes it into an attribution account.

4.6 Evaluator Base Model Analysis

To address RQ4, we vary only the evaluation agent's base LLM, comparing GPT-3.5-turbo, GPT-4o-mini, and GPT-4o while retaining the other components.

Table 4 shows that GPT-4o yields the highest downstream scores on all three metrics. Compared with GPT-3.5-turbo and GPT-4o-mini, it improves SMART-L by 4.5% and 7.6%, respectively, relative to each alternative's score. This indicates that evaluator choice influences the effectiveness of prompt optimization and the resulting reports' agreement with expert references.

Table 4: Effect of the evaluation agent's base LLM.

Eval. Base	BERTScore	SMART-L	ROUGE-L
GPT-3.5-turbo	0.8892	0.2701	0.3249
GPT-4o-mini	0.8789	0.2624	0.3344
GPT-4o	**0.9226**	**0.2823**	**0.3644**

The two alternatives exhibit different metric profiles: GPT-3.5-turbo performs better on BERTScore and SMART-L, whereas GPT-4o-mini leads on ROUGE-L. Evaluator feedback can therefore affect semantic and lexical alignment differently. Among the tested models, GPT-4o provides the most consistently effective feedback under these downstream metrics.

5 Conclusion

We introduce RERA, a multi-agent framework that integrates policy, macroeconomic, and financial evidence for real estate enterprise risk attribution. Its recognition, reasoning, and attribution stages connect external sources to firm-level outcomes. Textual gradients from expert-report evaluation refine the attribution prompt without changing model parameters. Experiments based on historical data from listed real estate companies show gains over general and finance LLMs on reference-based metrics, while human evaluation identifies room for improvement toward expert report quality. Future work will extend RERA to broader industries, longer time horizons, and more diverse risk scenarios.

REFERENCES

Reinald Kim Amplayo, Peter J. Liu, Yao Zhao, and Shashi Narayan. SMART: Sentences as basic units for text evaluation. *arXiv preprint arXiv:2208.01030*, 2022. URL https://arxiv.org/abs/2208.01030.

Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, Binyuan Hui, Luo Ji, Mei Li, Junyang Lin, Runji Lin, Dayiheng Liu, Gao Liu, Chengqiang Lu, Keming Lu, Jianxin Ma, Rui Men, Xingzhang Ren, Xuancheng Ren, Chuanqi Tan, Sinan Tan, Jianhong Tu, Peng Wang, Shijie Wang, Wei Wang, Shengguang Wu, Benfeng Xu, Jin Xu, An Yang, Hao Yang, Jian Yang, Shusheng Yang, Yang Yao, Bowen Yu, Hongyi Yuan, Zheng Yuan, Jianwei Zhang, Xingxuan Zhang, Yichang Zhang, Zhenru Zhang, Chang Zhou, Jingren Zhou, Xiaohuan Zhou, and Tianhang Zhu. Qwen technical report. *arXiv preprint arXiv:2309.16609*, 2023.

Yupeng Cao, Zhi Chen, Prashant Kumar, Qingyun Pei, Yangyang Yu, Haohang Li, Fabrizio Dimino, Lorenzo Ausiello, K.P. Subbalakshmi, and Papa Momar Ndiaye. Risklabs: Predicting financial risk using large language model based on multimodal and multi-sources data. In *2025 IEEE International Conference on Data Mining Workshops (ICDMW)*, pp. 901–909, 2025. doi: 10.1109/ICDMW69685.2025.00107.

Kaiji Chen and Yi Wen. The great housing boom of china. *American Economic Journal: Macroeconomics*, 9(2):73–114, April 2017. doi: 10.1257/mac.20140234. URL https://www.aeaweb.org/articles?id=10.1257/mac.20140234.

Rongda Chen, Jingjing Yu, Chenglu Jin, Xinyang Chen, Liu Yang, and Shuonan Zhang. Political connection and credit risk of real estate enterprises: evidence from stock market. *Humanities and Social Sciences Communications*, 11(1):1–12, December 2024a. doi: 10.1057/s41599-023-02522-z. URL https://ideas.repec.org/a/pal/palcom/v11y2024i1d10.1057_s41599-023-02522-z.html.

Weize Chen, Yusheng Su, Jingwei Zuo, Cheng Yang, Chenfei Yuan, Chi-Min Chan, Heyang Yu, Yaxi Lu, Yi-Hsin Hung, Chen Qian, Yujia Qin, Xin Cong, Ruobing Xie, Zhiyuan Liu, Maosong Sun, and Jie Zhou. Agentverse: Facilitating multi-agent collaboration and exploring emergent behaviors. In *International Conference on Learning Representations*, volume 2024, pp. 20094–20136, 2024b. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/578e65cdee35d00c708d4c64bce32971-Paper-Conference.pdf.

Yueyan Chen, Jianxin Xu, Shuyu Wang, and Sisi Xu. Economic environment uncertainty and financialization of real estate firms. *International Review of Economics & Finance*, 93:1104–1114, 2024c. ISSN 1059-0560. doi: https://doi.org/10.1016/j.iref.2024.05.011. URL https://www.sciencedirect.com/science/article/pii/S1059056024003095.

Jihye Choi, Jinsung Yoon, Long T. Le, Somesh Jha, and Tomas Pfister. Policybank: Evolving policy understanding for llm agents. *arXiv preprint arXiv:2604.15505*, 2026. URL https://arxiv.org/abs/2604.15505.

Gheorghe Comanici, Eric Bieber, Mike Schaekermann, et al. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*, 2025. URL https://arxiv.org/abs/2507.06261.

J. Delgadillo, J. Kinyua, and C. Mutigwe. FinSoSent: Advancing financial market sentiment analysis through pretrained large language models. *Big Data and Cognitive Computing*, pp. 87, 2024.

Qiu Deng, Rafael Alvarado, Fang Cheng, Lizeth Cuesta, Chun Wang, and Stefania Pinzón. Long-run mechanism for house price regulation in china: Real estate tax, monetary policy or macro-prudential policy? *Economic Analysis and Policy*, 77:174–186, 2022. doi: 10.1016/j.eap.2022.11.009.

H. Ding, Y. Li, J. Wang, et al. Large language model agent in financial trading: A survey, 2024. arXiv preprint arXiv:2408.06361.

David Gabauer and Rangan Gupta. Spillovers across macroeconomic, financial and real estate uncertainties: A time-varying approach. *Structural Change and Economic Dynamics*, 52:167–173, 2020. ISSN 0954-349X. doi: https://doi.org/10.1016/j.strueco.2019.09.009. URL https://www.sciencedirect.com/science/article/pii/S0954349X1930219X.

S. Gao, Y. Wen, M. Zhu, et al. Simulating financial market via large language model based agents. *arXiv preprint arXiv:2406.19966*, 2024.

M. Geerts, M. Reusens, B. Baesens, et al. On the performance of LLMs for real estate appraisal. In *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 195–211, Cham, 2025. Springer Nature Switzerland.

GLM Team, Aohan Zeng, Bin Xu, Bowen Wang, et al. ChatGLM: A family of large language models from GLM-130b to GLM-4 all tools. *arXiv preprint arXiv:2406.12793*, 2024. URL https://arxiv.org/abs/2406.12793.

A. Gotmare, N. S. Keskar, C. Xiong, and R. Socher. A closer look at deep learning heuristics: Learning rate restarts, warmup and distillation. In *Proceedings of the International Conference on Learning Representations*, 2019.

Feijiang Han, Hengtao Cui, Licheng Guo, Zhiyuan Lyu, and Zhiyuan Lyu. Read before you think: Mitigating llm comprehension failures with step-by-step reading. *arXiv (Cornell University)*, abs/2504.09402, 2025. doi: 10.48550/arxiv.2504.09402.

Yuzhi Hao and Danyang Xie. A multi-llm-agent-based framework for economic and public policy analysis. *arXiv preprint arXiv:2502.16879*, 2025. URL https://arxiv.org/abs/2502.16879.

S. Hong, X. Zheng, J. Chen, et al. MetaGPT: Meta programming for multi-agent collaborative framework, 2023. arXiv preprint arXiv:2308.00352.

Matteo Iacoviello. House prices, borrowing constraints, and monetary policy in the business cycle. *American Economic Review*, 95(3):739–764, June 2005. doi: 10.1257/0002828054201477. URL https://www.aeaweb.org/articles?id=10.1257/0002828054201477.

Song Jin, Shuqi Li, Shukun Zhang, and Rui Yan. FinRpt: Dataset, evaluation system and LLM-based multi-agent framework for equity research report generation. *arXiv preprint arXiv:2511.07322*, 2025. URL https://arxiv.org/abs/2511.07322.

A. Kim, M. Muhn, and V. Nikolaev. Financial statement analysis with large language models, 2024. arXiv preprint arXiv:2407.17866.

J. R. Landis and G. G. Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, 1977.

J. Li, X. Yuan, Q. Xie, et al. Homogeneity of corporate risk perceptions and systemic financial risk. *Humanities and Social Sciences Communications*, 2025.

Xiao-Lin Li, Lijuan Wang, and Dongmin Kong. Macro-prudential policy and systemic risk of real estate firms: Evidence from china. *Finance research letters*, 58:104518–104518, 2023. doi: 10.1016/j.frl.2023.104518.

Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pp. 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics. URL https://aclanthology.org/W04-1013/.

X. Y. Liu, G. Wang, H. Yang, et al. FinGPT: Democratizing internet-scale data for financial large language models. *arXiv preprint arXiv:2307.10485*, 2023.

Llama Team, AI at Meta. The Llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024. URL https://arxiv.org/abs/2407.21783.

X. Ma and H. Xie. Real estate policy regulation and corporate financial risk: China's three red lines policy. *Pacific Economic Review*, 30(1):46–87, 2025.

Dylan McNamara, Martin Smith, Zachary Williams, Sathya Gopalakrishnan, and Craig Landry. Policy and market forces delay real estate price declines on the us coast. *Nature communications*, 15(1):2209–2209, 2024. doi: 10.1038/s41467-024-46548-6.

L. J. Mena, V. Garcia, V. G. Felix, et al. Enhancing financial risk prediction with symbolic classifiers: Addressing class imbalance and the accuracy-interpretability trade-off. *Humanities and Social Sciences Communications*, 11(1):1–11, 2024.

J. W. M. Mwamba, J. C. Mba, and A. K. Kitenge. Navigating uncertainty in an emerging market: Data-centric portfolio strategies and systemic risk assessment in the johannesburg stock exchange. *International Journal of Financial Studies*, 13(1):32, 2025.

Ogonna Nneji, Chris Brooks, and Charles W.R. Ward. House price dynamics and their reaction to macroeconomic changes. *Economic Modelling*, 32:172–178, 2013. ISSN 0264-9993. doi: https://doi.org/10.1016/j.econmod.2013.02.007. URL https://www.sciencedirect.com/science/article/pii/S0264999313000576.

OpenAI. GPT-4o system card. *arXiv preprint arXiv:2410.21276*, 2024. URL https://arxiv.org/abs/2410.21276.

B. Paranjape, S. Lundberg, S. Singh, H. Hajishirzi, L. Zettlemoyer, and M. T. Ribeiro. ART: Automatic multi-step reasoning and tool-use for large language models. *arXiv preprint arXiv:2303.09014*, 2023.

D. Peskoff and B. M. Stewart. Credible without credit: Domain experts assess generative language models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 427–438, 2023.

Aske Plaat, Annie Wong, Suzan Verberne, Joost Broekens, Niki van Stein, and Thomas Back. Multi-step reasoning with large language models, a survey. *arXiv (Cornell University)*, 2024. doi: 10.48550/arxiv.2407.11511.

Reid Pryzant, Dan Iter, Jerry Li, Yin Lee, Chenguang Zhu, and Michael Zeng. Automatic prompt optimization with "gradient descent" and beam search. In Houda Bouamor, Juan Pino, and Kalika Bali (eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pp. 7957–7968, Singapore, December 2023. Association for Computational Linguistics. doi: 10.18653/v1/2023.emnlp-main.494. URL https://aclanthology.org/2023.emnlp-main.494/.

H. Shi, B. Zhang, and X. Nie. Analysis on the contribution of the new development of China's real estate industry to systemic financial risks: Taking august 2023 as an example. *Finance*, 14(3): 87695, 2024.

N. Shinn, F. Cassano, A. Gopinath, K. Narasimhan, and S. Yao. Reflexion: Language agents with verbal reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 36, 2023.

X. Tang, X. Wang, W. X. Zhao, S. Lu, Y. Li, and J.-R. Wen. Unleashing the potential of large language models as prompt optimizers: Analogical analysis with gradient-based model optimizers. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(24):25264–25272, 2025.

Z. Wang, E. Khatibi, and A. M. Rahmani. MedCoT-RAG: Causal chain-of-thought RAG for medical question answering, 2025. arXiv preprint arXiv:2508.15849.

J. Wei, X. Wang, D. Schuurmans, et al. Chain-of-thought prompting elicits reasoning in large language models. In *Advances in Neural Information Processing Systems*, volume 35, pp. 24824–24837, 2022.

S. Yao, D. Yu, J. Zhao, I. Shafran, T. L. Griffiths, Y. Cao, and K. Narasimhan. Retroformer: Retrospective large language agents with policy gradient optimization. *arXiv preprint arXiv:2308.02151*, 2023a.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models. *arXiv preprint arXiv:2210.03629*, 2023b. URL https://arxiv.org/abs/2210.03629.

Y. Yu, Z. Yao, H. Li, et al. FinCon: A synthesized LLM multi-agent system with conceptual verbal reinforcement for enhanced financial decision making. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

A. Zaremba and E. Demir. ChatGPT: Unlocking the future of NLP in finance, 2023. In Modern Finance, 93–98.

Kourosh Zarinjooei, Hasan Ghodrati, Hossein Panahian, and Hossein Jabbari. Identifying and prioritizing the factors affecting enterprise risk management implementation. *Iranian Journal of Finance*, 9(1):134–161, 2025. ISSN 2676-6337. doi: 10.61186/ijf.2024.448623.1463. URL https://www.ijfifsa.ir/article_210099.html.

C. Zhang, X. Liu, M. Jin, et al. When AI meets finance (StockAgent): Large language model-based stock trading in simulated real-world environments. *arXiv preprint arXiv:2407.18957*, 2024.

Hong Zhang, Linjun Li, Tiantian Chen, and Vera Li. Where will china's real estate market go under the economy's new normal? *Cities*, 55:42–48, 2016. ISSN 0264-2751. doi: https://doi.org/10.1016/j.cities.2016.03.014. URL https://www.sciencedirect.com/science/article/pii/S0264275116300555.

T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi. BERTScore: Evaluating text generation with BERT. In *International Conference on Learning Representations*, 2020.

Xuanyu Zhang and Qing Yang. Xuanyuan 2.0: A large chinese financial chat model with hundreds of billions parameters. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, CIKM '23, pp. 4435–4439, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400701245. doi: 10.1145/3583780.3615285. URL https://doi.org/10.1145/3583780.3615285.

W. Zhu, J. Liao, Z. Zhan, et al. Systematic review and future perspectives of financial distress prediction studies: Back to the principle of finance. *Journal of Tsinghua University (Science and Technology)*, 63(9):1467–1482, 2023.

A Dataset

This appendix describes the process of constructing the RERA dataset and its distribution.

A.1 Data Collection

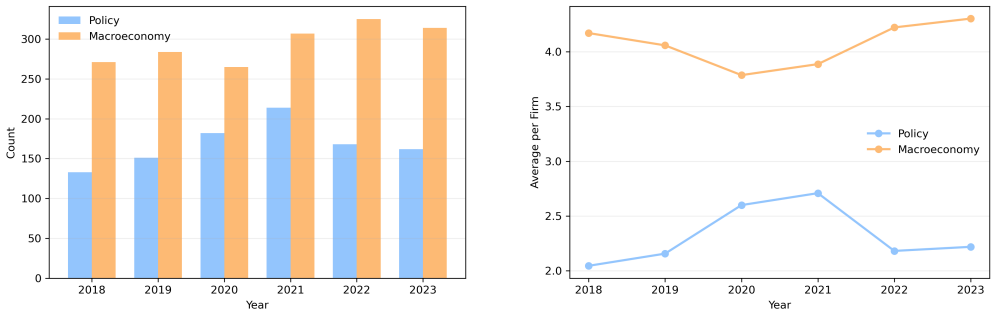
The data collection process begins with identifying listed real estate enterprises. We first selected 88 listed property companies and obtained their annual reports for the period 2018–2023. Furthermore, based on the information in the annual reports, we collected relevant policy data and macroeconomic data. Policy data may include both local and national policies. Macroeconomic data were then calibrated against the figures in the annual reports using officially published documents. Finally, we used the AKShare API to extract financial data for each company, including indicators such as total assets, revenue, cash flow, and debt-to-asset ratio. The detailed content of each data type and data source of the data collection is present in Table 5.

Data Type	Introduction	Data Source
Policy data	National and local real estate policies, such as mortgage rules, purchase restrictions, and financing regulations	Shanghai Stock Exchange Announcement Official Policy Releases
Macroeconomic data	Macroeconomic variables and macro-level real estate data, such as GDP growth, money supply, credit conditions, real estate investment, sales area, sales value, and housing price index	Shanghai Stock Exchange announcement Official Statistics
Financial data	Enterprise financial and operational indicators, such as total assets, revenue, operating cash flow, debt-to-asset ratio, and other key financial metrics	AKShare API

Table 5: The introduction of data contents and data sources.

A.2 Data Distribution

Figure 3 reports the annual total and per-firm average counts of policy and macroeconomic problem sources. Between 2018 and 2023, on average, each enterprise faced approximately four macroeconomic risk factors and two to three policy-related risk factors per year. Figure 4 shows the asset-scale



(a) Annual total counts.

(b) Average counts per firm.

Figure 3: Distribution of enterprise risk sources from 2018 to 2023.

distribution of the covered enterprises by comparing their total assets in 2018 and 2023. The sample includes both large-scale enterprises with assets at the trillion-level and smaller listed firms. This broad asset range helps ensure that the dataset is not limited to a single enterprise scale. Changes in assets also reflect heterogeneous firm development: many enterprises expanded their asset scale, while several firms experienced asset contraction, providing varied firm-level conditions for risk attribution.

B Agent Prompt Design

This appendix provides a detailed account of the representative prompt templates employed by various agents within the RERA system, with the aim of elucidating our approach to encoding different

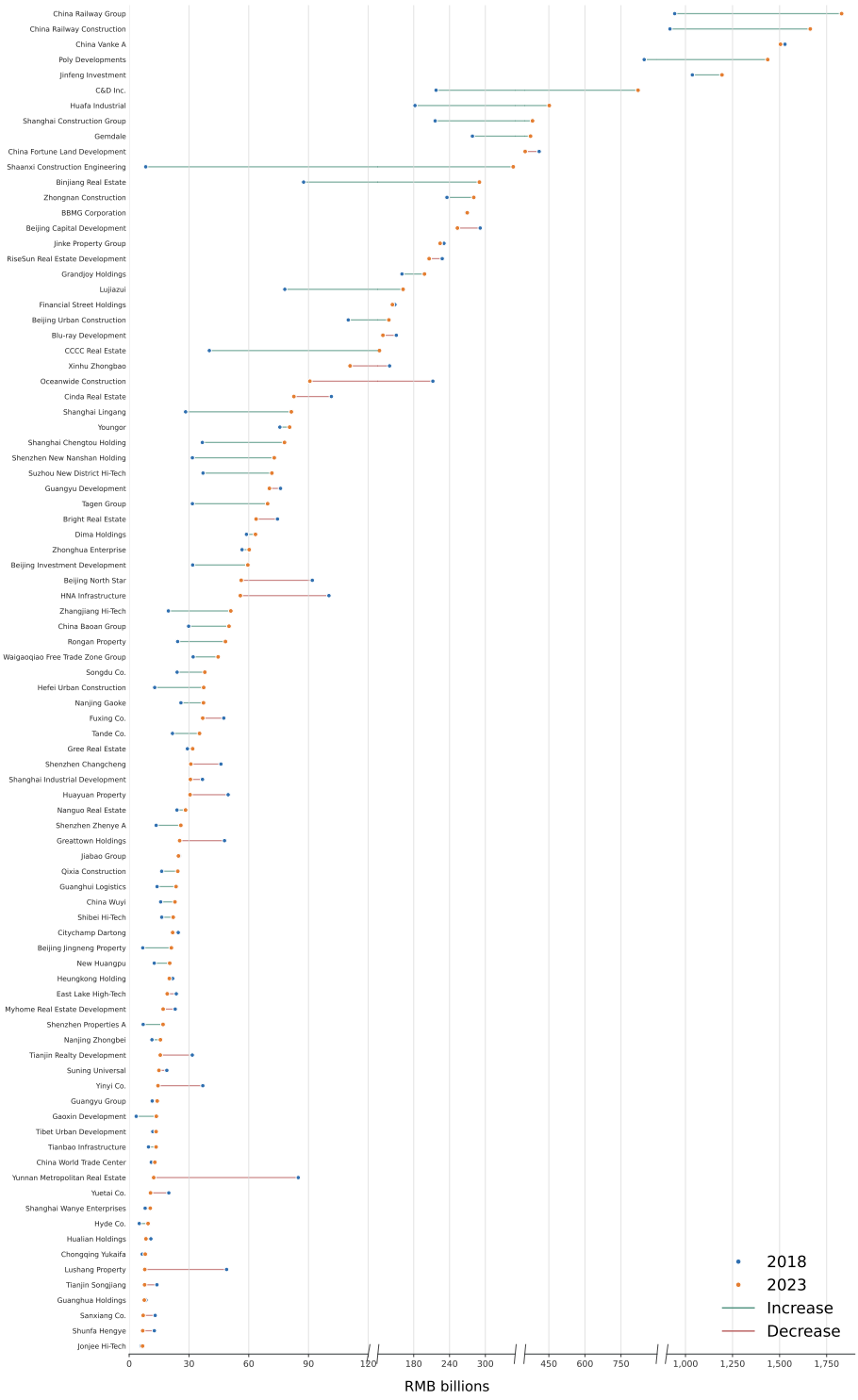


Figure 4: Distribution of enterprise total assets in 2018 and 2023. Blue and orange points represent total assets for the respective years, while connecting segments indicate whether total assets increased or decreased over the period.

functionalities and expert reasoning paradigms. Each template has been carefully designed to guide

LLMs in performing specific analytical and reasoning tasks, thereby ensuring that their behavioral characteristics align with expectations whilst also generating structured outputs.

C Example Risk Attribution Report

This section presents a qualitative risk-attribution example under the "Housing Is for Living, Not for Speculation" policy. For the same Vanke-centered case, Table 6 compares how RERA, GPT-4o, and TongYi-Finance-14B-chat organize the policy source, transmission channels, and firm-level risk outcomes.

The comparison highlights RERA's intended behavior. RERA keeps the policy source explicit and traces it through credit, mortgage, land, and tax channels to cash flow, revenue, and investment return indicators. GPT-4o identifies liquidity and solvency pressure, but it mixes broad market statistics with firm-level claims without preserving a complete risk causal chain. TongYi-Finance-14B-chat structures the policy effect into sales, financing, and land-acquisition risks, yet its explanation remains closer to an industry-level policy analysis and provides limited Vanke-specific numerical grounding.

D Prompt Revision Process

This section describes how the attribution agent prompt is optimized during the revision process. The initial instruction follows the three-stage design in the agent prompt template: recognition stage, reasoning Stage, and attribution stage. During optimization, feedback from generated reports is converted into stage-specific prompt updates while preserving the same sentence frames and making only word-level additions or deletions. The recognition stage is strengthened to select supported sources and preserve exact source names. The reasoning stage is strengthened to connect each source to a concrete business channel and firm-level indicator. The attribution stage is strengthened to write only evidence-grounded risk outcomes. After these revisions, the prompt provides clearer constraints for source selection, transmission-chain reasoning, and numerical grounding, leading to better attribution reports. Table 7 shows the changes in prompts across each stage.

Policy Agent Prompt Template

Profile:
You are a real estate policy extraction agent. Your responsibility is to convert official policy text into structured policy observations for enterprise risk attribution.

Input:
Firm metadata: {firm name}
Policy evidence: {policy document}, {annual-report policy excerpt}

Instruction:
Preserve the original policy keyword and summarize the core policy content, including issuer, time, regional scope, target behavior, and affected market mechanism. Distinguish policy evidence from macroeconomic statistics and firm financial facts. Merge semantically equivalent policy descriptions only when they refer to the same policy signal, while keeping the most specific source keyword. Explain the likely affected enterprise business area, such as sales, financing, land acquisition, inventory, pricing, presale funds, rental housing, or property management.

Output Format:
Please provide your refined analysis using the structure below:
[Policy Keyword]: <1–3 keywords appearing in the policy text>
[Policy Source]: <The authority issuing the policy>
[Policy Abstract]: <An overview of the policy's content>
[Policy Imapct]: <The potential impact of this policy on which areas of business for real estate enterprises>

Figure 5: Policy Agent prompt template.

Macroeconomic Agent Prompt Template

Profile:
You are a macroeconomic interpretation agent. Your role is to convert macro and market-level indicators into structured observations for enterprise risk attribution.

Input:
Firm metadata: {firm name}
Macro evidence: {macroeconomic indicators}, {industry/market statistics}

Instruction:
Extract only macroeconomic or market-level signals with explicit values, units, periods, and directions. Focus on housing sales area/value, development investment, new starts, land market, price expectations, credit conditions, and real estate loan growth. For each signal, summarize the phenomenon, classify the direction as expansion, contraction, tightening, easing, recovery, or volatility when supported, and explain the business channel through demand, financing, land acquisition, inventory turnover, cash collection, pricing, or investment pace.

Output Format:
Please provide your refined analysis using the structure below:
[Economic Phenomenon]: <Summarising current economic phenomena by analysing trends in key indicators>
[Supporting Indicator]:<Provide the indicators from the data supporting the conclusion as evidence>
[Change Direction]: <Analyse the trends in the various indicators>
[Potential Impact]: <Potential business risks arising from this phenomenon>

Figure 6: Macroeconomic Agent prompt template.

Financial Agent Prompt Template

Profile:
You are a real estate financial analysis agent. Your role is to convert enterprise financial data into structured observations for risk attribution.

Input:
Firm metadata: {firm name}
Financial evidence: {quarterly/annual financial statements}, {annual report analysis}, {prepared financial data}

Instruction:
Extract enterprise-level financial signals with explicit values, periods, and trends. Analyze asset changes, liability structure, cash-flow condition, and profit performance. Focus on asset/liability growth, debt-to-asset ratio, inventories, receivables, operating cash flow, financing cash flow, revenue, net profit, and gross margin. Summarize abnormal changes and the business situations they imply, such as solvency pressure, liquidity constraint, inventory turnover risk, financing burden, profit compression, delivery pressure, or dependence on non-recurring gains. Preserve reported figures and do not add policy or macro causes.

Output Format:
Please provide your refined analysis using the structure below:
[Financial Indicator]: <Provide key financial indicators and values from the data>
[Change Direction]: <Analyze year or quarter trends and abnormal changes>
[Business Situation]: <Summarize operating conditions reflected by indicators>
[Potential Risk]: <Potential financial risks arising from the business situation>

Figure 7: Financial Agent prompt template.

Attribution Agent Prompt Template

Profile:
You are a real estate enterprise risk attribution agent. Your role is to integrate analyst observations into traceable source-channel-risk chains.

Input:
Firm metadata: {firm name}
Analyst evidence: {policy observations}, {macroeconomic observations}, {financial observations}

Instruction:
Recognition Stage:
You need to identify the specific policies and macroeconomic factors affecting the company's operations from the analyzed data by selecting relevant source items, preserving exact keywords, and classifying each item as policy or macroeconomic evidence.
Reasoning Stage:
Establishing the transmission channels through which policy measures and macroeconomic changes affect companies' specific business operations requires linking each source to sales, financing, land acquisition, inventory turnover, cash collection, pricing, delivery, or cost pressure, and checking the link against financial observations.
Attribution Stage:
You need to establish a causal chain to clarify how policy and macroeconomic changes give rise to business risks, forming a source to risk outcome path and marking unsupported links as not attributable.

Output Format:
Please provide your refined analysis using the structure below:
[Source Type]: <Policy or macroeconomic factor>
[Specific Source]: <Exact policy keyword, macro indicator, or market signal>
[Content Summary]: <Summarize the source content>
[Impact Chain]: <A comprehensive explanation of the risk transmission chain from the source of risk to financial indicators>
[Risk Outcome]: <Analysis of the ultimate impact of the risk on the enterprise>

Figure 8: Attribution Agent prompt template.

Evaluation Agent Prompt Template

Profile:
You are an evaluator of real estate risk attribution reports. Compare the model output with the expert report and evidence, then return scores and diagnostic feedback for prompt optimization.

Input:
Model output: {generated attribution report}
Ground truth: {expert attribution report}
Evidence: {policy and macro observations}, {financial observations}

Instruction:
Evaluate whether the report selects supported policy or macroeconomic sources, keeps the correct impact direction, and links each source to concrete business operations and financial evidence. Logic Consistency (LC) checks whether the causal chain is complete and self-consistent. Factual Alignment (FA) checks whether the selected source, attribution direction, and variable relationships match the expert report and input evidence. Quantitative Reasonableness (QR) checks whether numerical claims or magnitude descriptions are close to the expert report or real data in direction and scale.

Output Format:
Please provide scores for the three dimensions, along with the basis for those scores, using the structure below:
[Logic Consistency (LC)]: <0-100 score and reason>
[Factual Alignment (FA)]: <0-100 score and reason>
[Quantitative Reasonableness (QR)]: <0-100 score and reason>

Figure 9: Evaluation Agent prompt template.

Table 6: Comparison of risk attribution report excerpts under the "Housing Is for Living, Not for Speculation" policy.

Model	Risk Attribution Analysis
RERA	**Source Type:** Policy **Specific Source:** "Housing Is for Living, Not for Speculation" **Content Summary:** The Politburo meeting emphasized a firm commitment to resolving real estate market issues by adhering to city-specific policies, promoting supply–demand balance, guiding market expectations, rectifying market order, and curbing excessive housing price increases. A long-term mechanism for the stable and healthy development of the real estate market was proposed. On the financial side, authorities continuously tightened real estate credit policies, strictly controlling loan growth for developers and strengthening personal mortgage regulation by raising down payment ratios for second homes, thereby suppressing speculative demand. In land policy, local governments adjusted residential land supply structures and increased the proportion of land for rental housing. Tax reforms aligned with this principle were also promoted, including pilot property tax measures in some cities to increase holding costs for multiple-home owners and restrain speculative ownership. **Impact Chain:** The policy affects corporate cash flows through financial, land, and tax channels. Credit tightening and higher down payments directly suppressed speculative purchases, extending Vanke's project inventory turnover cycle from an average of 8–10 months to 12–15 months. Operating cash inflows slowed significantly, with net operating cash flow decreasing by 35.53% year-on-year. Mortgage tightening reduced new housing demand, delaying the recovery of accumulated presale funds and forcing price cuts in some cities, which compressed profit margins and reduced operating revenue by 10.05%. Increased rental land supply constrained land development options; given Vanke's capital focus on commodity housing, some land projects were abandoned, reducing investment returns by 20%. Although property tax reforms suppressed speculation, Vanke's stable community mitigated long-term cash flow risks. **Risk Outcome:** Through financial tightening, land structure adjustment, and tax reforms, the policy significantly increased Vanke's cash flow risk. Slower sales recovery and higher expenditures intensified liquidity pressure. Net operating cash flow declined by 35.53% year-on-year, operating revenue fell by 10.05%, and investment returns decreased by 20%. Under dual pressure from sluggish inflows and rigid outflows, funding chain risks further intensified.
GPT-4o	Affected by the "Housing Is for Living, Not for Speculation" policy, Vanke's sales growth slowed. National commercial housing sales area increased by only 1.3%, while revenue growth was about 22%. Inventory surged by 52% to approximately RMB 750 billion, advance receipts dropped sharply to below RMB 300 million, and cash holdings fell to RMB 18.8 billion, accounting for only 1.2% of total assets. Total liabilities rose to about RMB 1.29 trillion, with a debt ratio of 84.6%. Financial expenses increased by nearly 190% year-on-year, indicating significant sales pressure, severe inventory accumulation, tight cash flows, and elevated solvency risks due to high leverage.
TongYi-Finance-14B-chat	Under the overarching principle that 'houses are for living in, not for speculation', regulatory policies, which involve city-specific measures to stabilise land prices, house prices and market expectations, have directly given rise to three key risks for enterprises. On the sales front, these city-specific measures have led to severe market polarisation across different cities, with sales volumes declining in third- and fourth-tier cities, whilst the pace of demand release in key cities has been disrupted by expectations of regulatory intervention, leaving enterprises facing operational pressures amid this market polarisation. On the financing front, industry regulations have imposed strict oversight on various financing channels, with the growth rate of property loans continuing to slow. The difficulty for enterprises to secure financing has increased across the board, thereby creating risks of cash flow constraints. On the land acquisition front, the objective of stabilising land prices has reshaped the valuation logic of the land market. Market activity has cooled significantly in the second half of the year, with an increase in transactions at reserve prices and unsold plots. Consequently, the stability of enterprises' land reserve acquisition and their cost control are now subject to risks arising from fluctuations in the pace and cost of land acquisition.

Revision Agent Prompt Template

Profile:
You are a prompt revision agent for textual-gradient prompt optimization. Your goal is to improve the attribution agent prompt using previous prompts, Logic Consistency(LC), Factual Alignment(FA), Quantitative Reasonableness(QR) scores, and evaluator feedback.

Input:
Current prompt: {current attribution prompt}
Feedback: {LC/FA/QR scores}, {main weakness}
History and edit budget: {previous prompt-score records}, {allowed word-level edit distance}

Instruction:
Carefully analyse the previous prompts and their scores, and draft a new attribution prompt with the aim of achieving a higher evaluation score. Use the feedback on low-scoring prompts to strengthen various capabilities, such as evidence binding, checking the direction of impact, refining the source-channel-risk chain, and numerical calculations. Draw on consistent phrasing from historically high-scoring prompts, but ensure that all modifications remain within the specified scope and avoid introducing redundant or repetitive content.

Output Format:
Change at most {allowed word-level edit distance} words relative to the current prompt.
Only output the revised prompt itself in the following form:
Candidate Prompt:<revised attribution agent prompt set>

Figure 10: Revision Agent prompt template.

Recognition Stage	Reasoning Stage	Attribution Stage
Step 0. You need to identify the specific policies and macroeconomic factors affecting the company's operations from the analyzed data by selecting relevant source items, preserving exact keywords, and classifying each item as policy or macroeconomic evidence.	**Step 0.** Establishing the transmission channels through which policy measures and macroeconomic changes affect companies' specific business operations requires linking each source to sales, financing, land acquisition, inventory turnover, cash collection, pricing, delivery, or cost pressure, and checking the link against financial observations.	**Step 0.** You need to establish a causal chain to clarify how policy and macroeconomic changes give rise to business risks, forming a source to risk outcome path and marking unsupported links as not attributable.
Step 5. You need to identify the specific policies and macroeconomic factors explicitly supported by analyst evidence and affecting the target company's operations from the analyzed data by selecting relevant source items, preserving exact keywords and original source phrases, and classifying each item as policy or macroeconomic evidence.	**Step 5.** Establishing the transmission channels through which retained policy measures and macroeconomic changes affect companies' specific business operations requires linking each retained source to sales, financing, land acquisition, inventory turnover, cash collection, pricing, delivery, or cost pressure, explaining the link in words, and checking the link and effect direction against financial observations.	**Step 5.** You need to establish a causal chain to clarify how selected policy and macroeconomic changes give rise to business risks, forming a supported source to channel to risk outcome path and marking delayed, offset, or unsupported links as not attributable.
Step 10. You need to identify the specific policies and macroeconomic factors consistently supported by analyst evidence and affecting the target company's operations from the analyzed data by selecting relevant source items, preserving exact keywords and original source phrases, merging duplicated source names, and classifying each item as policy or macroeconomic evidence.	**Step 10.** Establishing the transmission channels through which retained policy measures and macroeconomic changes affect companies' specific business operations requires linking each retained source to sales, financing, land acquisition, inventory turnover, cash collection, pricing, delivery, cost pressure, cash flow, revenue, liabilities, margins, or land-reserve cost, explaining the link in words, and checking the link and effect direction against firm-level financial observations.	**Step 10.** You need to establish a causal chain to clarify how selected policy and macroeconomic changes give rise to business risks, forming a supported source to channel to financial-indicator to risk outcome path and marking delayed, offset, unsupported, or unquantified links as not attributable.
Step 15. You need to identify the specific policies and macroeconomic factors consistently supported by analyst evidence and affecting the target company's firm-level operations from the analyzed data by selecting only relevant source items, preserving exact keywords and source names, merging duplicated source names, and classifying each item consistently as policy or macroeconomic evidence.	**Step 15.** Establishing the complete transmission channels through which retained policy measures and macroeconomic changes affect companies' specific business operations requires linking each retained source to sales, financing, land acquisition, inventory turnover, cash collection, pricing, delivery, cost pressure, cash flow, revenue, liabilities, margins, or land-reserve cost, explaining the source-channel-indicator link in words, and checking the link and effect direction against firm-level evidence.	**Step 15.** You need to establish a causal chain to clarify how selected policy and macroeconomic changes give rise to evidence-grounded business risks, forming a supported source to channel to financial-indicator to quantified risk outcome path and marking delayed, offset, generic, unsupported, or unquantified links as not attributable in the report.

Table 7: Attribution prompt optimization trajectory.