# Knowledge-augmented Financial Market Analysis and Report Generation

Yuemin Chen[1,2,3], Feifan Wu[1,2,3], Jingwei Wang[2], Hao Qian[2], Ziqi Liu[2], Zhiqiang Zhang[2], Jun Zhou[2], Meng Wang[1]

[1]College of Design and Innovation, Tongji University, China

[2]Ant Group, China

[3]School of Computer Science and Engineering, Southeast University, China

## ABSTRACT

Crafting a convincing financial market analysis report necessitates a wealth of market information and the expertise of financial analysts, posing a highly challenging task. While large language models (LLMs) have enabled the automated generation of financial market analysis text, they still face issues such as hallucinations, errors in financial knowledge, and insufficient capability to reason about complex financial problems, which limits the quality of the generation. To tackle these shortcomings, we propose a novel task and a retrieval-augmented framework grounded in a financial knowledge graph (FKG). The proposed framework is compatible with commonly used instruction-tuning methods. Experiments demonstrate that our framework, coupled with a small-scale language model fine-tuned with instructions, can significantly enhance the logical consistency and quality of the generated analysis texts, outperforming both large-scale language models and other retrieval-augmented baselines.

## INTRODUCTION

We introduce Financial Market Analysis Generation (FMAG) as a task focused on creating logical and high-quality analytical reports using market data.

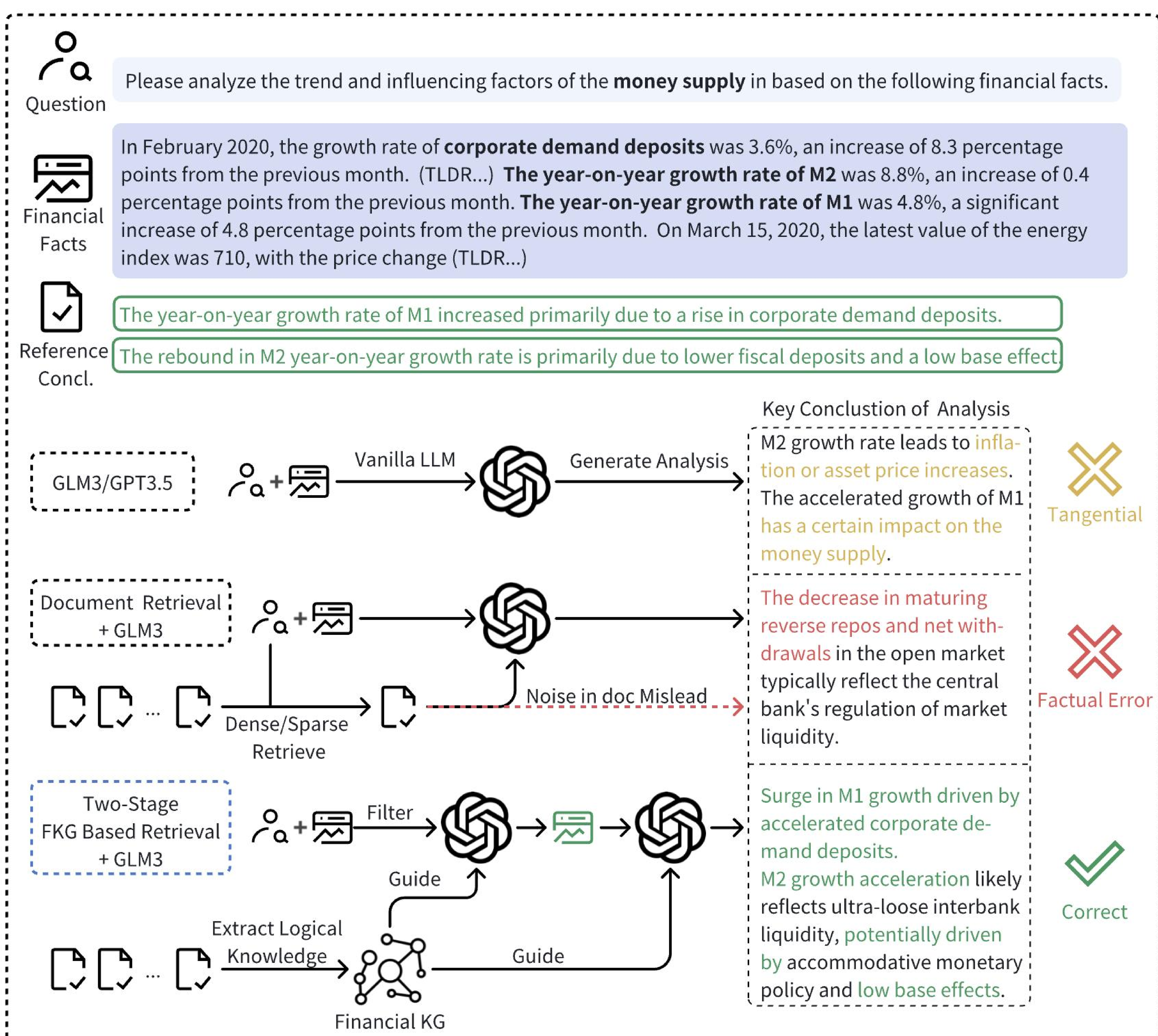


Figure 1: A comparison of FMAG between our method and other baselines.

Challenges in Financial Market Analysis Generatio:

- Crafting high-quality financial market analysis is complex, requiring vast market data and expert knowledge.
- Large language models (LLMs) have made strides in automating text generation for financial analysis but face issues like hallucinations, knowledge errors, and limited reasoning abilities, affecting the quality and reliability of outputs, see Fig. 1.

Proposed Solution:

- Financial Knowledge Graph (FKG): A comprehensive FKG is constructed using LLMs to capture complex relationships in financial data.
- Clustering-based Triple Retrieval: A retrieval strategy, enabling efficient retrieval from FKG automatically extracted
- Two-stage RAG: Combines information selection and subsequent reasoning based on FKG insights.

## TASK DESCRIPTION

FMAG is designed to produce analytical texts by reasoning with financial market data, including financial indicators, trends, and policy impacts. To simulate real-world FMKG, we developed a benchmark focused on bond market analysis.

We structure FMAG as a Question Answering task with explanations.

Each instance in FMAG consists of:

- Q: A user question.
- F: Relevant financial facts.
- A: The generated analysis, detailing reasoning steps and conclusions.

Objective:

Given Q and F, FMAG aims to estimate reasoning steps and produce the analysis text A, formulated as the probability P(A|Q, F).

## PROPOSED FRAMEWORK

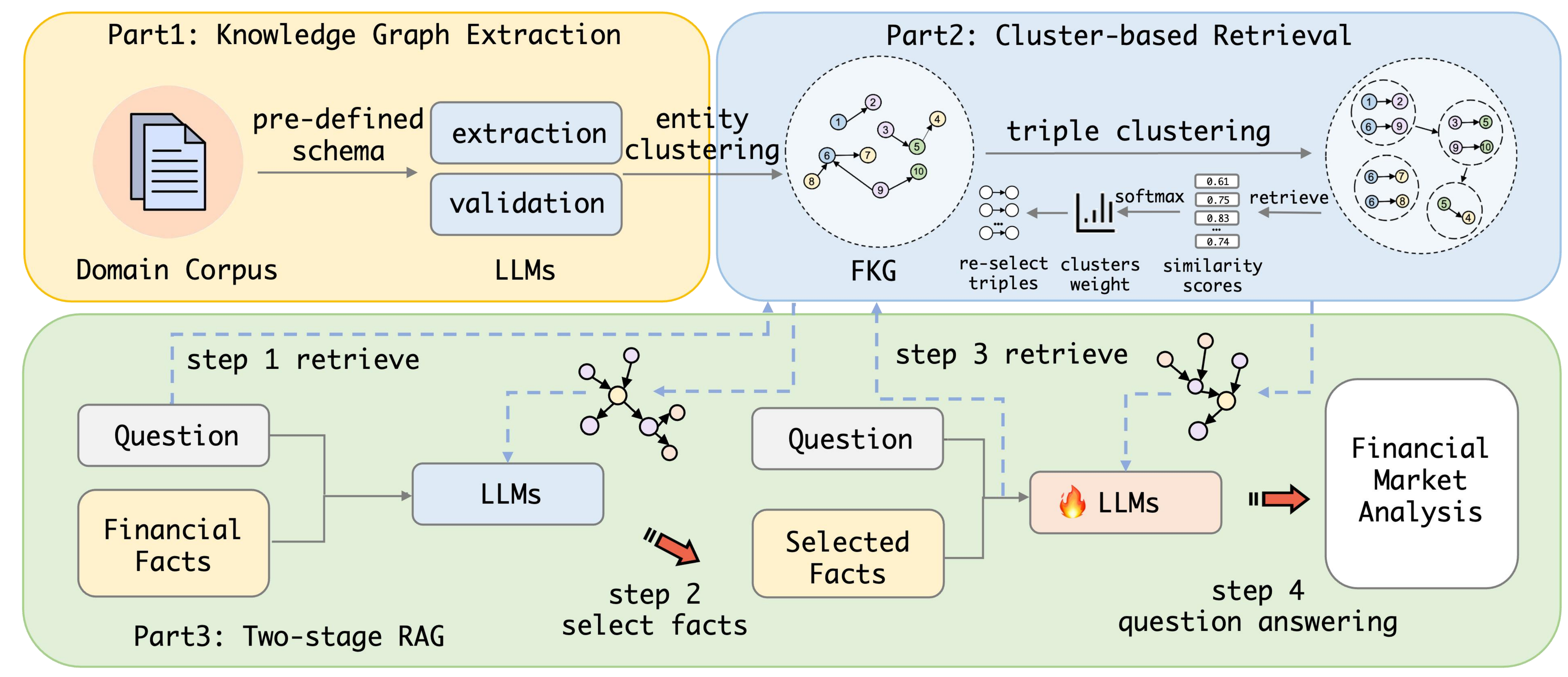


Figure 2: The overall framework for our Two-stage FKG-based Retrieval (TFR).

We introduce a two-stage FKG-based retrieval-augmented framework shown in Fig. 2. First, we build a FKG via prompting LLMs. Second, we propose a clusters-based retrieval method to facilitate the retrieval of triples. Thirdly, we propose a two-stage RAG method, in which the KG serves as guidance to conduct initial information selection in the first stage and reasoning in the second stage.

## RESULTS

Main Result:

- GLM3-turbo + TFR and GLM3-6b (SFT with FKG) + TFR achieved the top scores in GLM4-score Concl, highlighting TFR's advantage in conclusion accuracy.
- GLM3-6b (SFT with FKG) + TFR also led in GLM4-Score Text and RougeL, proving the effectiveness of combining SFT with TFR across metrics.
- Results of GLM3-6b (SFT w/o FKG) suggesting that SFT mainly enhances language style alignment with reference text rather than improving reasoning in conclusions.

Table 2: **The results for different models on our benchmark.** GLM4-Score Concl. denotes the consistency score of the generated text and reference conclusion. GLM4-Score Text denotes the consistency score of generated text and reference text. TFR denotes our Two-Stage FKG based retrieval method. The highest score is denoted in **bold**, and the second-highest score is underlined.

| Metric | GLM4-Score | | BERT Score | | | RougeL | | |
|---|---|---|---|---|---|---|---|---|
| Model | Concl. | Text | P | R | F1 | P | R | F1 |
| GPT3.5-turbo | 2.8625 | 2.4502 | 0.6309 | 0.7341 | 0.677 | 0.4672 | 0.4244 | 0.3952 |
| GLM3-turbo | 2.8247 | 2.5464 | 0.6265 | 0.7351 | 0.675 | 0.4057 | 0.4709 | 0.3891 |
| GLM3-turbo + BM25 Retrieve | 2.9661 | 2.539 | 0.6232 | 0.732 | 0.6719 | 0.3322 | 0.4716 | 0.3377 |
| GLM3-turbo + Dense Retrieve | 3.0761 | 2.737 | 0.6336 | 0.7515 | 0.6862 | 0.3678 | 0.5281 | 0.382 |
| GLM3-turbo + Triples Retrieve | 3.2136 | 2.9492 | 0.6371 | 0.7332 | 0.6803 | 0.4333 | 0.4742 | 0.4094 |
| GLM3-turbo + TFR | **3.3254** | 2.9966 | 0.6328 | 0.7267 | 0.6751 | 0.3441 | 0.4728 | 0.3504 |
| GLM3-6b | 2.7424 | 2.3932 | 0.6579 | 0.7331 | 0.6907 | 0.3048 | 0.5162 | 0.3127 |
| GLM3-6b (SFT w/o FKG) | 2.9424 | 3.4373 | **0.8546** | **0.7878** | **0.8178** | 0.6184 | 0.707 | 0.5911 |
| GLM3-6b (SFT with FKG) | 3.0949 | 3.4712 | 0.8536 | 0.7629 | 0.8034 | 0.6788 | 0.5775 | 0.5708 |
| GLM3-6b (SFT with FKG) + TFR | 3.2203 | **3.5593** | 0.8393 | 0.7728 | 0.8023 | **0.7438** | **0.6384** | **0.6474** |

Ablation Study:

- The TFR model improves GLM4 scores, with greater gains for stronger models.
- Triple integration in SFT enhances small models' use of retrieved info and boosts accuracy, especially in conclusions.
- Models lacking triple integration in SFT show performance drops.

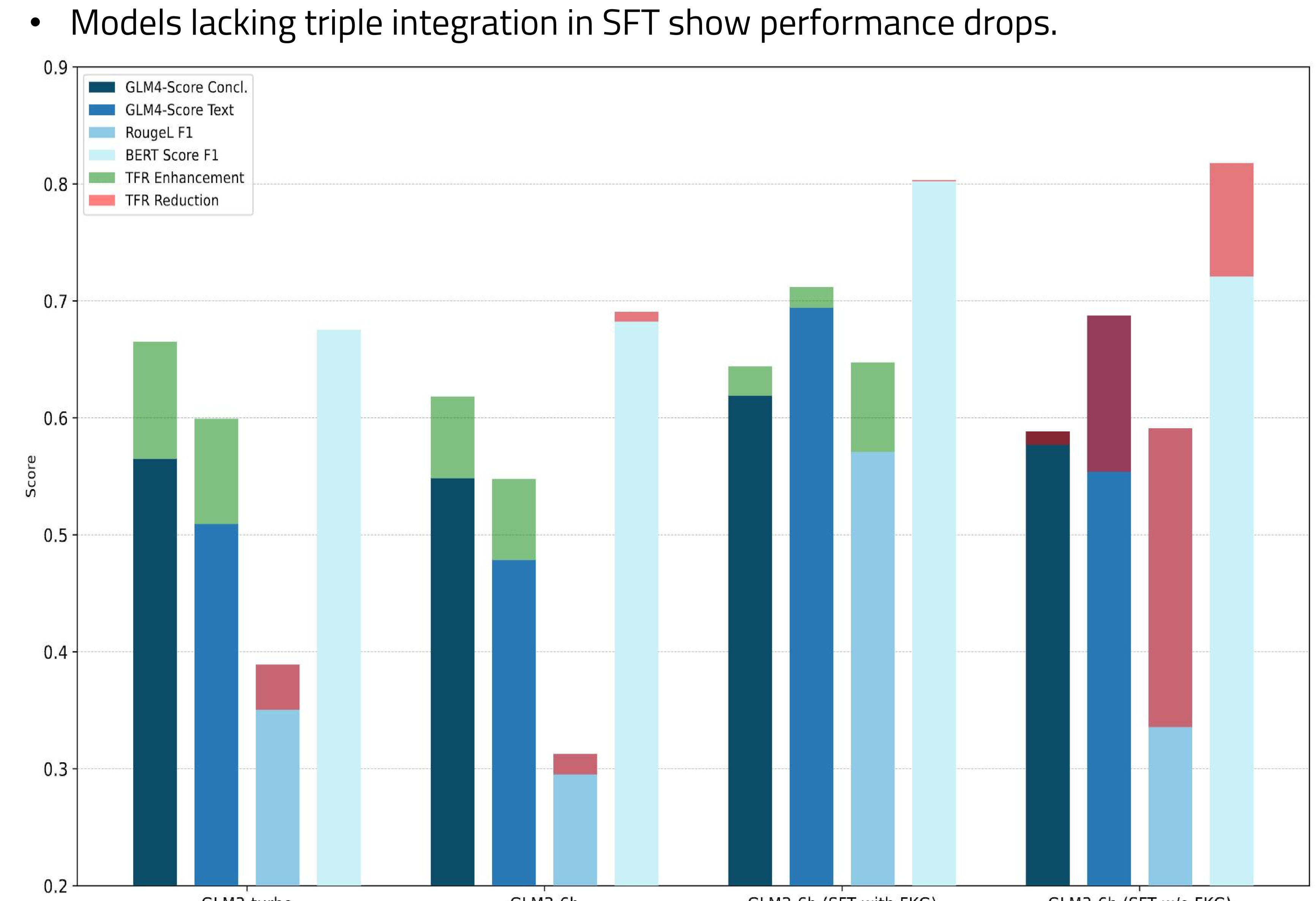


Figure 3: Comparative analysis of model performance on our benchmark with different components

## ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China (U23B2057, 62176185,62276063), the Natural Science Foundation of jiangsu Province (BK20221457), and the Ant Group.