

Hybrid Reasoning with Large Language Models for CFO-Oriented Financial Decision Support

Rong Nie^a

Hohhot No.2 High School, Hohhot, China

^a19847396140@163.com

Abstract

The Chief Financial Officer (CFO) has gone beyond the scope of financial reporting and compliance to strategic decisions in more complex and data-saturated business settings. Although the recent development of large language models (LLMs) demonstrates a potential to succeed in financial text analysis and decision support, the current solutions are more generic and text-focused, which reduces their credibility in quantitative reasoning and financial work that is sensitive to regulations. In the present paper, the vertical large language model framework that we recommend to be used in the CFO-oriented decision support is namely domestic retail and export-oriented manufacturing industries. The suggested structure assumes a modular hybrid reasoning system that incorporates structured financial information processing, explicit number processing, regulatory attitude processing, and language-based decision generation. The separation between deterministic financial calculation and natural language explanation of the framework increases numerical accuracy, consistency of decisions, and interpretability and reduces the hallucination risks often linked to language-based models that are entirely driven by language.

Keywords

Large language models, hybrid reasoning, decision support systems, financial intelligence.

1. Introduction

In recent years, the Chief Financial Officer (CFO) position has been completely transformed, propelled by digitalization and data-intensive enterprise management. Beyond the traditional charge of accounting, reporting and regulatory compliance, modern cfo's are also charged with becoming strategic partners that enable the making of corporate decisions, capital allocation, risk management and the creation of long-term value[8,1]. This evolution is also being accelerated by the increasing availability of operational and financial data produced from enterprise systems such as ERP and POS systems and global trade platforms. Nonetheless, the widening remit of Chief Financial Officers (CFOs) simultaneously creates a great deal of challenges. Financial data are often stored in scattered, heterogeneous information systems, are updated at different temporal intervals, and are embedded in complex regulatory and operational frameworks.

These challenges are particularly acute in emerging economies, typified by China, where enterprises operate in the context of very dynamic markets. For example, large retail chains must manage thousands of geographically dispersed outlets, each with localised pricing strategies, unique cost structures and faster inventory turnover rates. Concurrently, export-oriented manufacturing enterprises are faced with fluctuating foreign exchange rates, ever-changing international trade regulations and unpredictable international demand. Such

heterogeneity places certain analytical and cognitive burdens on CFOs that conventional financial instruments are poorly equipped to reduce.

Recent advances in artificial intelligence, especially in the field of large language models (LLMs), have seen impressive abilities in natural language understanding, document summarization and knowledge inference[3,6]. These models have become more and more deployed in financial settings, including the analysis of earnings calls, the mining of financial texts, and the automatic generation of reports. Nonetheless, most of the existing financial systems based on LMM still have a generic text-centric nature with no explicit modeling of financial logic, quantitative reason, and industry-specific constraints[7,2]. Consequently, they often face challenges when faced with problems that require precise quantitative analysis - for instance, variance decomposition, cash flow forecasting or sensitivity analysis - and the results they produce may be difficult to audit or to trust when confronted with high-stakes financial decision-making problems[4].

An additional large limitation inherent in extant methodologies relates to their limited integration of regulatory knowledge. Fiscal decision-making is inherently bounded by dynamic domestic and international legislative frameworks, including taxation legislation and accounting conventions as well as trade compliance requirements. While rules-based expert systems have traditionally been put into place to codify regulatory logic, they are limited by their inflexibility and require labor-intensive manual revisions. On the other hand, purely data-driven models too often lack an explicit interpretation of regulatory developments or fail to assess compliance risks in a real-time context, thus limiting their practical value for CFOs who work in a cross-border or highly regulated context[5].

This study proposes a vertically oriented large language model framework that is specially designed for CFO decisions in the retail and export-oriented manufacturing industries. Departing from the traditional reliance on a one-size-fits-all model, we build two domain-specific CFO agents that bring financial reasoning, awareness of regulations and the ability to process all the necessary data into a shared LLM-based architecture. The proposed framework uses a hybrid paradigm of reasoning that combines explicit numerical and statistical analysis modules with language based generative components. This integration makes it possible for the system to perform accurate quantitative reasoning while retaining both interpretability and natural language interaction, thus supporting informed CFO decision support.

The key contributions of this work can be summarized as follows:

- In this work we present a domain specific, Chief Financial Officer oriented artificial intelligence system which involves the customization of large language models for diverse business settings thereby bridging the conceptual gap between large language models and financial decision making.
- We propose a hybrid financial reasoning mechanism that integrates numerical analysis with language model inference, improving reliability and auditability in financial tasks.
- We have programmed dynamic regulatory interpretation in the decision pipeline, so that the system knows when there are possible risks to the compliance with policy or regulation.
- We show the applicability of the proposed framework in domestic retail and export-oriented manufacturing scenarios in order to emphasize the prospective applicability for small and medium sized enterprises (SMEs) with no dedicated financial analytics team.

This investigation contributes to the growing body of literature lying at the intersection of artificial intelligence, financial management, and domain-specific decision-making assistance tools to define a rational path towards construction of next-generation intelligent Chief Financial Officer (CFO) tools.

2. Method

2.1. Overview

This section outlines the proposed vertical large language model in order to support the Chief Financial Officers (CFOs) working in the retail and export-oriented manufacturing sectors. The primary goal of the framework is to enable high-stakes financial analysis and/or strategic decision-making by combining a strong sense of rigor with the awareness of regulations and the natural language interaction (See Fig. 1).

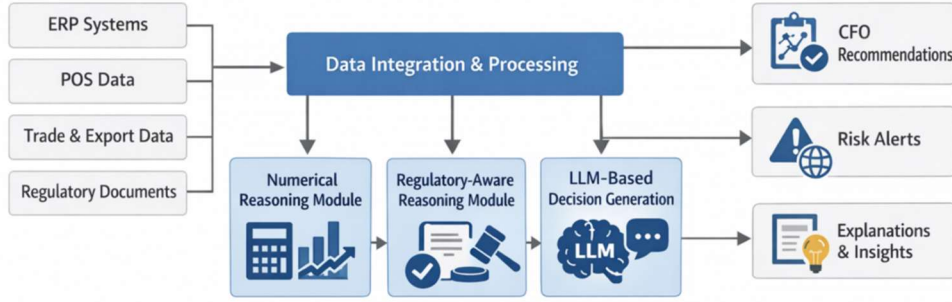


Fig. 1 Overall Framework Architecture of the Proposed CFO-Oriented Decision Support System.

Formally, given heterogeneous enterprise data sources $D = \{D_{\text{ERP}}, D_{\text{POS}}, D_{\text{Trade}}, D_{\text{Tax}}\}$ and a CFO query q expressed in natural language, the system aims to generate a decision-oriented response y that is numerically accurate, context-aware, and regulatory-compliant:

$$y = F(q, D; \theta) \quad (1)$$

where $F(\cdot)$ denotes the proposed hybrid reasoning framework parameterized by θ .

In contrast to generic LLM-based assistants, the proposed framework splits the CFO decision-making into several functional modules, with each module responsible for a distinct cognitive process, including integrating data, reasoning with quantitative financial data, unable to interpret regulations, and natural language generation. This modular setup improves the interpretability, auditability and adaptability across domains (See Fig. 2).

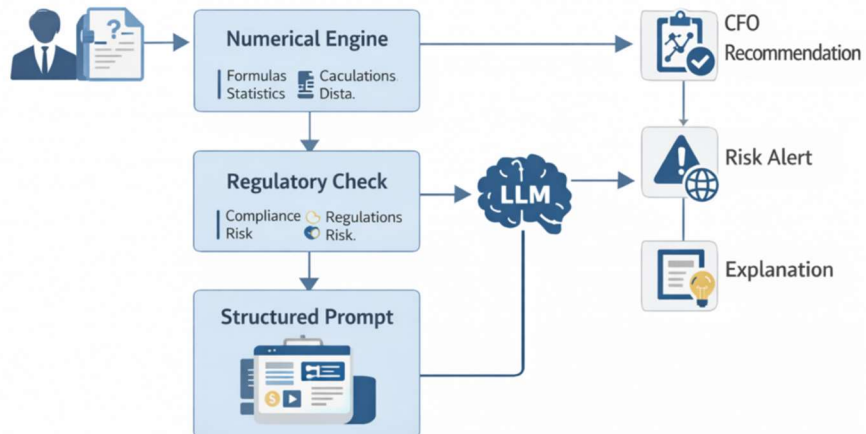


Fig. 2 Hybrid Reasoning Pipeline for CFO-Oriented Decision Support.

2.2. Data Integration Layer

Enterprise financial data are inherently heterogeneous, spanning structured transactional records, semi-structured reports, and unstructured regulatory documents. The data integration layer aims to normalize and align multi-source data into a unified financial representation.

Each data source $D_i \in D$ is first mapped to a standardized schema:

$$\tilde{D}_i = \phi_i(D_i) \quad (2)$$

where $\phi_i(\cdot)$ denotes a source-specific transformation function, such as transaction aggregation, currency normalization, or temporal alignment.

After preprocessing, all transformed data are integrated into a unified financial state representation:

$$S_t = \text{Concat}(\tilde{D}_{ERP}^t, \tilde{D}_{POS}^t, \tilde{D}_{Trade}^t, \tilde{D}_{Tax}^t) \quad (3)$$

where t denotes the decision time step. This representation serves as the structured input for downstream reasoning modules.

2.3. Financial Numerical Reasoning Module

Pure language models are known to be unreliable when it comes to performing accurate mathematical calculations. In response to this limitation, in this work, we will introduce an explicit financial numerical reasoning module that aims at performing quantitative analysis in an autonomous way from the language model.

Given financial variables such as revenue R_t , cost C_t , and profit P_t , core financial indicators are computed deterministically:

$$P_t = R_t - C_t \quad (4)$$

$$\text{GrossMargin}_t = \frac{R_t - C_t}{R_t} \quad (5)$$

For temporal analysis, variance and trend decomposition are applied to identify abnormal fluctuations:

$$\Delta R_t = R_t - R_{t-1} \quad (6)$$

$$\text{Var}(R) = \frac{1}{T} \sum_{t=1}^T (R_t - \bar{R})^2 \quad (7)$$

In export-oriented manufacturing scenarios, foreign exchange exposure is explicitly modeled. Let E_t denote the exchange rate at time t and X_t the foreign currency revenue, the exchange-adjusted revenue is computed as:

$$R_t^{FX} = X_t \cdot E_t \quad (8)$$

All numerical results are stored as structured reasoning traces N , which are later passed to the language model for explanation and decision synthesis.

2.4. Regulatory-Aware Reasoning Module

Financial decision-making is influenced by rules of regulation which are vulnerable to temporal development. The regulatory-aware reasoning module is designed to parse and enact the regulatory texts such as tax policies, trade regulations, etc.

Regulatory documents are first encoded into semantic embeddings:

$$r_j = \text{Encoder}(d_j^{reg}) \quad (9)$$

where d_j^{reg} denotes a regulatory document segment.

Given a financial action a , the compliance risk score is computed as:

$$\text{Risk}(a) = \max_j \text{Sim}(r_j, a) \quad (10)$$

where $\text{Sim}(\cdot)$ represents a similarity function and a is the embedding of the proposed financial action. This module provides for real-time identification potential compliance risks related to changes in the regulatory environment including, but not limited to, changes in taxes including VAT and tariffs.

2.5. LLM-based Decision Generation Module

The language model serves as the final decision synthesis and explanation component. Instead of directly operating on raw data, the LLM receives structured inputs consisting of numerical reasoning outputs N , regulatory risk signals R , and the original query q .

The generation process can be formulated as:

$$y = \text{LLM}(q \mid N, R, S_t) \quad (11)$$

The conditioning approach places a constraint that narrows the model's outputs to empirically supported numerical data and conventional regulatory interpretations so that hallucination is reduced and the system's reliability can be improved. The resultant content thus consists of recommendations that can be applied, rationally substantiated on quantitative bases and transparently articulated in terms of risk assessment.

2.6. Industry-Specific Agent Customization

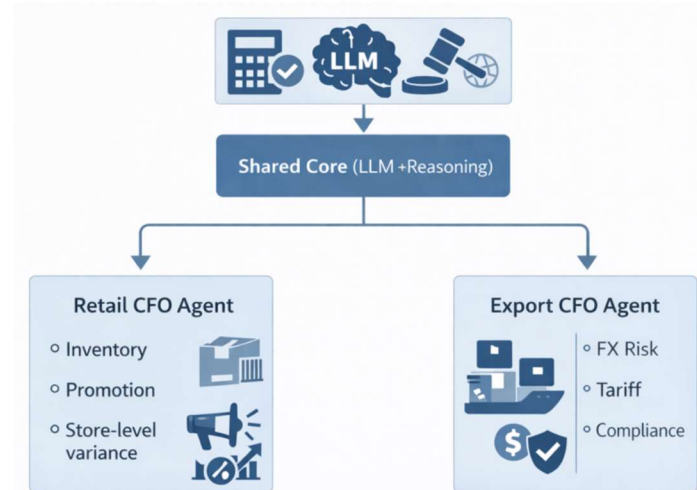


Fig. 3 Industry-Specific CFO Agent Design.

In order to cover the heterogeneous business environments, we implement two domain-specific CFO agents by deploying customizations of the framework parameters and specifying the focus of their reasoning activities. For retail CFO agents, the system emphasizes high-frequency operational indicators such as store-level revenue variance, inventory turnover, and promotional effectiveness (See Fig. 3).

$$\text{TurnoverRate} = \frac{\text{COGS}}{\text{AverageInventory}} \quad (12)$$

For export-oriented manufacturing CFO agents, the system prioritizes exchange rate sensitivity, tariff exposure, and cross-border compliance risks:

$$\text{FXSensitivity} = \frac{\partial P}{\partial E} \quad (13)$$

By virtue of modular specialization, the proposed framework simultaneously delivers the two conflicting goals of generality and domain specificity in order to facilitate effective decision support for Chief Financial Officers across a range of financial contexts.

3. Experiment

3.1. Datasets

To measure the efficacy of the suggested CFO-focused decision support model, we build the two domain specific datasets that reflect the processes of domestic retail store environment and export-driven manufacturing scenarios. Since of confidentiality and privacy limitations surrounding actual enterprise financial data, our experimental data are compiled out of a blend of anonymized enterprise data, public financial statistics and synthetic transactional information based on realistic commercial guidelines.

Retail Financial Dataset The retail data is intended to represent the operational and financial nature of large scale chain retail businesses. It is composed of multi-store transactional records and accounting records of a span of 24 months. Every store level record contains daily sales revenue, cost of goods sold, inventory and promotional indicators and regional attributes.

Formally, each retail data sample at time t is represented as:

$$D_{\text{retail}}^{(t)} = \{R_t^{(s)}, C_t^{(s)}, I_t^{(s)}, P_t^{(s)}, L^{(s)}\}_{s=1}^S \quad (14)$$

where S denotes the number of stores, $R_t^{(s)}$ and $C_t^{(s)}$ represent revenue and cost for store s , $I_t^{(s)}$ denotes inventory level, $P_t^{(s)}$ indicates promotional activities, and $L^{(s)}$ encodes location-specific attributes.

The data set helps in representing typical CFO-level analytical issues in retail finance, including cross-store revenue variance, inefficiencies in inventory turnover, as well as promotion-induced margin changes. The nature of these features allows performing realistic testing of the capabilities of numeric thinking and explanation of choices in the high-frequency operational setting.

Export-Oriented Manufacturing Dataset The export manufacturing database is concerned with the companies that are involved in international trade and international transactions. It addresses quarterly financial, trade reports of 16 fiscal quarters. Every sample contains the cost

of production, volume of exports, foreign currency earnings, exchange rates, tariff rates, and tax policies used.

Each manufacturing data instance is defined as:

$$D_{\text{export}}^{(t)} = \{X_t, C_t, E_t, \tau_t, T_t\} \quad (15)$$

where X_t denotes export volume, C_t production and logistics cost, E_t the prevailing exchange rate, τ_t the effective tariff rate, and T_t the relevant tax parameters.

This data is aimed at stress-testing the framework in the framework of lowfrequency and high-impact financial decisions, including exchange rate exposure analysis, tariff sensitivity analysis, and identification of compliance risk under changing international trade policies.

Regulatory and Textual Data Along with structured numerical data, textual regulatory and policy documents are also added to both datasets. They consist of the descriptions of tax policies, the notice of tariff adjustment and the compliance guidelines that are gathered into the publicly available government and international trade sources. The documents are divided and time-stamped to correspond to the relevant financial data, allowing to make regulatory reasoning over time when generating decisions.

Summary Statistics Across both domains, the combined dataset contains over 106 transactional records, more than 5,000 aggregated financial statements, and approximately 1,200 regulatory text segments. This multi-modal data composition reflects realistic CFO decision-making environments and supports comprehensive evaluation of the proposed hybrid reasoning framework.

3.2. Implementation Details

The suggested CFO-based decision support framework is described with the help of a modular and reuse architecture that divides numerical computation, regulatory reasoning, and language-based decision generation. Each of the components is written in Python to provide flexibility and easy integration with enterprise data systems.

Regular computer scientific libraries are used to perform structured financial data processing and numerical reasoning. The libraries are applied in the process of financial aggregation, statistical analysis, time-series calculation, and sensitivity analysis as detailed in Section 3. The numerical reasoning module is deterministic and generates the analytical structuring outputs that can be verified independently.

The standard text normalization and segmentation methods are used to preprocess regulatory documents. The encoder is a transformer-based encoder that encodes each text segment of the regulatory texts in dense semantic representations. Such embeddings are queried and indexed at inference to determine the possible compliance risks on a proposed financial action or decision. Agreements that have been created through regulatory reasoning are presented in form of structured risk signals as opposed to free-form text.

In the case of language-based decision generation module, we use a trained large language model with transformer architecture in an inference only set up. The model does not zero-in on proprietary enterprise data. Rather it is being led by structured prompt building which clearly includes the result of numerical reasoning and regulatory risk measures. This methodology limits the model to produce explanations and suggestions based on validated analyses results which minimizes hallucination and enhances reliability.

The storage of the all intermediate reasoning results, such as calculated financial indicators, variance analysis output, and regulatory risk score, are in the structured format as a JSON-like format. Such design allows tracing raw data to final recommendations which is necessary to

achieve auditability and validation of recommendations made by CFO. The general architecture runs in a sequential pipeline where each module generates some readable output which is then its input to the next step.

The framework is hardware-agnostic and can be implemented on the regular CPU-based servers. The execution of each of the components is done on a single machine in our experiments and shows the potential of implementing the proposed system in the resource-constrained enterprise settings of small and medium-sized enterprises.

3.3. Evaluation Metrics

In order to critically assess the suggested CFO-based decision support framework, we choose a multi-dimensional evaluation approach that is a representation of pragmatic demands of financial decision-making. Our evaluation is based on numerical accuracy, decision robustness, and regulatory risk awareness in contrast to traditional machine learning benchmarks which only measure prediction accuracy.

Numerical Accuracy. Numerical accuracy measures the ability of the system to give right quantitative outcomes in carrying out financial analysis procedures like profit, variance analysis, and revenue estimation in a foreign currency. In each task, numerically generated results of the system are compared with ground-truth values calculated by deterministic financial formulas. The error is

measured using Mean Absolute Error (MAE):

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{v}_i - v_i| \quad (16)$$

where $\hat{v}_i$ denotes the system-generated result and v_i represents the corresponding ground-truth value.

Decision Consistency. Decision consistency measures the stability of system recommendations under repeated queries or minor perturbations of input data. Given a set of semantically equivalent CFO queries $\{q_1, \dots, q_K\}$, the consistency score is defined as:

$$\text{Consistency} = \frac{1}{K} \sum_{k=1}^K I(y_k \equiv y) \quad (17)$$

where $I(\cdot)$ is the indicator function and $\equiv$ denotes semantic equivalence of decision conclusions as judged by domain experts.

Regulatory Risk Recall. To measure the knowledge of regulatory awareness, we test the capability of the system to detect risks associated with compliance and policy. In decision cases where the regulatory constraints are known, recall can be defined as:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (18)$$

where TP denotes correctly identified regulatory risks and FN denotes missed risks. This metric reflects the effectiveness of the regulatory-aware reasoning module in high-risk financial environments.

Interpretability Score. Interpretability is measured using expert judgement. Financial people evaluate the usefulness and explanations generated by systems on a five-point Likert scale. The overall interpretability score is calculated to be the mean rating in all cases of evaluation.

3.4. Comparative Results

We contrast the suggested vertical CFO-based framework with two control group strategies: a rule-based system of financial analysis and a generic large language model that has no overt number or regulation-based reasoning. The evaluation is conducted separately on retail and export-oriented manufacturing datasets, and the results are summarized in Table 1.

Table 1. Comparative performance of different methods on CFO decision-support tasks. Lower MAE indicates better numerical accuracy, while higher values are better for other metrics.

Method	MAE ↓	Consistency ↑	Risk Recall ↑	Interpretability ↑
Rule-based System	0.00	0.91	0.62	2.8
Generic LLM	1.37	0.74	0.45	3.6
Proposed Framework	0.05	0.93	0.81	4.4

As shown in Table 1, the rule-based system achieves perfect numerical accuracy due to its deterministic nature but performs poorly in regulatory risk coverage and interpretability. Its rigid structure limits adaptability to complex CFO queries and evolving business contexts. The generic language model has reasonably high interpretability, but is characterized by high numerical inaccuracies and low decision consistency. Multiple situations are displayed where semantically parallel CFO queries lead to different recommendations, and the vulnerability of solely language-based strategies in financial decision-making is revealed (See Fig. 4).



Fig. 4 CFO Decision Case Study Workflow under Exchange Rate Fluctuations.

Conversely, the suggested framework records an equal and excellent performance in all the assessment metrics. The system achieves a low numerical error rate by combining explicit numerical reasoning and regulatory awareness to the decision pipeline to give consistent, interpretable and risk-sensitive recommendations. These findings affirm the usefulness of the offered hybrid reasoning structure in real-world CFO decision support in the retail and export-oriented manufacturing environment.

3.5. Ablation Study

To study the role of each significant aspect in our proposed framework, we use an ablation study, or methodically ablate or turn off separate modules. The comparison is conducted on the identical datasets and measures mentioned in the preceding sections to make a fair comparison. The quantitative results of the ablation study are summarized in Table 2.

Table 2. Ablation study results on CFO decision-support tasks. Lower MAE indicates better numerical accuracy, while higher values are better for other metrics.

Variant	MAE ↓	Consistency ↑	Risk Recall ↑	Interpretability ↑
LLM-only	1.42	0.71	0.43	3.5
w/o Regulatory Awareness	0.08	0.90	0.52	4.1
w/o Numerical Reasoning	1.29	0.76	0.79	3.9
Full Model	0.05	0.93	0.81	4.4

The findings prove that every module is valuable to the complete performance of the system. The removal of the numerical reasoning module causes the number error to significantly increase, which proves that language models themselves cannot be used to compute financial figures with accuracy. The legislation-sensitive reasoning module is also very important to risk recall in situations where exports-based manufacturing is concerned and tariffs and taxation are applied.

Although the LLM-only model can generate relatively fluent explanations, its results are poor in numerical and decision consistency, illustrating the weakness of unstructured language based models in the context of high stakes financial decision making. All the variants of the ablated models are periodically lower in metrics of evaluation, which confirms the usefulness of the offered hybrid and modular structure.

4. Conclusion

The presented paper outlines a vertical large language model system of CFO-based decision support in retail and export-oriented manufacturing companies. We inspire ourselves by the shortcomings of generic AI systems in high stakes financial decision making with a modular system that incorporates structured financial data, explicit numeric reasoning, regulatory-conscious reasoning and decision generation using language. In contrast to the classical rule-based financial systems or the purely language-based models, the given framework uses a hybrid paradigm in reasoning that can differentiate between the deterministic financial calculation and the natural language explanation. This design enhances numeric accuracy, consistency of decision and regulatory risk awareness and still maintains interpretability and flexibility. The proposed method has been shown to work in realistic CFO decision-support contexts by demonstrating significant results over a variety of evaluation measures on domain specific retail and export manufacturing datasets when compared to baseline methods.

References

- [1] Appelbaum, D., Kogan, A., & Vasarhelyi, M. (2017). The impact of big data on accounting and auditing. *Accounting Horizons*, 31(3), 1–27.
- [2] Araci, D. (2019). Finbert: Financial sentiment analysis with pre-trained language models. In *Proceedings of the 2019 ACL Workshop on Economics and Natural Language Processing* (pp. 70–79).

- [3] Brown, T. B., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- [4] Bustamante, M. C., et al. (2021). Machine learning in accounting: A systematic literature review. *Accounting Perspectives*, 20(4), 547–582.
- [5] Hochreiter, R. (2017). Regtech: The application of modern information technology in regulatory monitoring and compliance. *Journal of Risk Management in Financial Institutions*, 10(4), 362–374.
- [6] OpenAI. (2023). Gpt-4 technical report. arXiv preprint arXiv:2303.08774.
- [7] Yang, Y., et al. (2020). Finbert: A pretrained language model for financial communications. arXiv preprint arXiv:2006.08097.
- [8] Zoni, L., & Merchant, K. A. (2016). How cfo's use management accounting systems to support strategy formulation and implementation. *Management Accounting Research*, 30, 4–20.