

# FinRobot: An Open-Source AI Agent Platform for Financial Applications using Large Language Models

Hongyang (Bruce) Yang<sup>1</sup>, Boyu Zhang<sup>1</sup>, Neng Wang<sup>1</sup>, Cheng Guo<sup>1</sup>,  
 Xiaoli Zhang<sup>1</sup>, Likun Lin<sup>1,2</sup>, Junlin (Jason) Wang<sup>1,4</sup>, Tianyu Zhou<sup>1</sup>, Mao Guan<sup>1</sup>,  
 Runjia (Luna) Zhang<sup>1,4</sup>, Christina Dan Wang<sup>1,3\*</sup>;

<sup>1</sup>AI4Finance Foundation; <sup>2</sup>Columbia University;

<sup>3</sup>Shanghai Frontiers Science Center of Artificial Intelligence and Deep Learning,  
 NYU Shanghai; Business Division, NYU Shanghai;

<sup>4</sup>Shanghai AI-Finance School ECNU;  
 contact@ai4finance.org;

## Abstract

As financial institutions and professionals increasingly incorporate Large Language Models (LLMs) into their workflows, substantial barriers, including proprietary data and specialized knowledge, persist between the finance sector and the AI community. These challenges impede the AI community’s ability to enhance financial tasks effectively. Acknowledging financial analysis’s critical role, we aim to devise financial-specialized LLM-based toolchains and democratize access to them through open-source initiatives, promoting wider AI adoption in financial decision-making.

In this paper, we introduce FinRobot, a novel open-source AI agent platform supporting multiple financially specialized AI agents, each powered by LLM. Specifically, the platform consists of four major layers: 1) the Financial AI Agents layer that formulates Financial Chain-of-Thought (CoT) by breaking sophisticated financial problems down into logical sequences; 2) the Financial LLM Algorithms layer dynamically configures appropriate model application strategies for specific tasks; 3) the LLMOps and DataOps layer produces accurate models by applying training/fine-tuning techniques and using task-relevant data; 4) the Multi-source LLM Foundation Models layer that integrates various LLMs and enables the above layers to access them directly. Finally, FinRobot provides hands-on for both professional-grade analysts and laypersons to utilize powerful AI techniques for advanced financial analysis. We open-source FinRobot at <https://github.com/AI4Finance-Foundation/FinRobot>.

## 1 Introduction

Financial analysis typically involves interpreting market trends, predicting economic outcomes, and providing in-

vestment strategies [Greenwald *et al.*, 2004; Penman, 2010; Berman and Knight, 2013]. Financial analysis precedes financial decisions. Suppose we were to distinguish analysis performed in the equity space into two categories. Then, there is fundamental analysis [Abarbanell and Bushee, 1997] that studies the companies and gives valuations, while technical analysis studies market action to forecast future price trends [Murphy, 1999].

The analysis is grounded in data. As the digital revolution progresses, more data becomes accessible for analysis. Given the ever-growing volume and complexity of the data involved in the process, financial professionals increasingly resort to algorithmic and machine intelligence for data processing [Awotunde *et al.*, 2021]. Artificial intelligence (AI) [Cao, 2022] has revolutionized financial analysis by automating tasks typically performed by human analysts, such as sentiment analysis [Medhat *et al.*, 2014; Sohangir *et al.*, 2018; Huang *et al.*, 2023; Zhang *et al.*, 2023b] and market prediction [Henrique *et al.*, 2019; Nabipour *et al.*, 2020; Kumar *et al.*, 2022; Jiang, 2021]. Traditional AI models were initially limited to straightforward, single-task operations. However, advancements in LLMs [Brown *et al.*, 2020; Wu *et al.*, 2023b; Yang *et al.*, 2023] and computing power have significantly enhanced AI’s capabilities. LLMs, with their extensive parameter space and diverse training datasets, are adept at assimilating global knowledge, understanding contexts, and executing complex logical reasoning. This expansion in capabilities not only allows managing multiple tasks simultaneously but also extends utility in the financial sector for more complex analysis and decision-making.

Building on the foundation of LLMs, AI agents have evolved to employ these models for an array of sophisticated functionalities, including planning, memory management, and tool usage, which facilitate intelligent and autonomous actions in real-world scenarios [Xi *et al.*, 2023; Gao *et al.*, 2023; Wang *et al.*, 2023c; Durante *et al.*, 2024]. These capabilities allow AI agents to execute complex financial analyses and operations that traditionally required extensive human intervention and expertise. Recent applications of AI agents, such as FinAgent [Zhang *et al.*, 2024] for trading and FinMem [Yu *et al.*, 2023c] for financial decision-

\*Corresponding author

making, underscore the growing reliance on these technologies to perform sophisticated financial operations. Despite their advances, several critical challenges remain unaddressed by these solutions:

- **Transparency Improvement:** In what ways can AI Agent-driven financial analysis platforms improve transparency in their decision-making processes to build trust among users?
- **Global Markets Adaptation:** How can AI agents effectively adapt to global stock markets’ multilingual and multicultural nuances to ensure comprehensive and accurate market analysis?
- **Model Diversity:** What are the limitations of relying solely on a single LLM architecture, such as GPT-4, for complex financial analysis, and how can diversifying model architectures enhance performance?
- **Real-Time Data Processing:** How can AI agent platforms efficiently manage and route vast amounts of financial data to ensure timely and precise financial analysis?

Recognizing these challenges, this paper introduces FinRobot, an innovative open-source AI agent platform that harnesses multi-source LLMs for diverse financial tasks. FinRobot refines financial workflows from data processing to strategy implementation, broadening access to advanced analytical tools and enhancing scalability and transparency. By utilizing diverse LLM architectures, it tailors solutions to global market needs, ensuring precise adaptation and performance optimization. The platform’s multi-source integration architecture facilitates the selection of accurate LLMs for specific tasks, and its real-time data processing pipeline assures timely financial analysis.

Our contributions are significant and address previous shortcomings in the field:

- **Comprehensive Financial AI Agent Framework:** FinRobot provides a holistic framework for developing financial AI agents that can perform a wide range of tasks from market forecasting to financial document analysis. Notably, FinRobot is the first AI agent platform dedicated to financial analysis, marking a significant advancement in the application of AI in finance.
- **Multi-Source LLM Integration:** FinRobot introduces a novel Smart Scheduler mechanism that allows for the seamless integration of multi-source LLMs, enabling the platform to leverage the strengths of various state-of-the-art LLMs and select the most suitable ones for specific financial tasks. This adaptability is crucial for handling the complexities of global financial markets and multilingual data.
- **Open-Source Platform for Financial AI:** By adopting an open-source approach, FinRobot fosters broader collaboration and accelerates innovation within the financial AI community. This openness not only promotes transparency but also allows for continuous enhancements and adaptations, which are essential in the rapidly evolving financial sector.

The remainder of this paper is organized as follows. Section 2 briefly reviews related work. In Section 3, we describe the four layers in FinRobot. In Section 4, we present the financial Chain-of-Thought (CoT) Prompting technique.

In Section 5, we present the performance evaluation for two demo applications. Section 6 concludes this work and points out directions for future work.

## 2 Related Work

### 2.1 LLM-Based AI Agents

AI Agents, powered by large language models (LLMs), are intelligent entities capable of perceiving environments, making decisions, and executing actions independently. Recent advancements in AI have led to the proliferation of LLM-based AI agents across multiple domains, significantly impacting areas ranging from creative industries to technical fields [Sumers *et al.*, 2023; Park *et al.*, 2023; Zeng *et al.*, 2023]. For example, MusicAgent [Yu *et al.*, 2023a] enhances music composition, MedAgents [Tang *et al.*, 2023; Thirunavukarasu *et al.*, 2023] improves medical diagnostics, Data Analytics Agent [Qiao *et al.*, 2023] provides insights from big data, Programmer Agent [Rasheed *et al.*, 2024] automates coding tasks, and ResearchAgent generates research ideas [Baek *et al.*, 2024]. These applications illustrate the versatility of LLMs and set the stage for their pivotal role in advancing financial technology.

### 2.2 Financial AI Agents

In the financial sector, AI agent-driven systems like FinAgent [Zhang *et al.*, 2024] and FinMem [Yu *et al.*, 2023c] leverage LLMs to transform trading strategies by utilizing real-time market data for informed decision-making. However, the focus on metrics such as backtesting and individual stock returns often overshadows the need for robust, process-oriented approaches. Effective trading agents should not only perform well statistically but also enhance operational workflows, automating tasks from strategy execution to order placement with minimal human intervention. This shift from performance to process orientation can lead to more sustainable and adaptable financial technologies.

### 2.3 Open-Source AI Agent Frameworks

The development of open-source AI agent frameworks has been instrumental in democratizing access to advanced AI technologies. Prominent platforms and frameworks such as AutoGPT [Gravitas, 2023], AutoGen [Wu *et al.*, 2023a], MetaGPT [Hong *et al.*, 2023], HuggingGPT [Shen *et al.*, 2024], ChatDev [Qian *et al.*, 2023], Dify [LangGenius, 2024] and Voyager [Wang *et al.*, 2023b] facilitate collaborative enhancements and innovation by allowing a global community of developers to contribute and refine AI models. Given the growing need for sophisticated financial tools, the emergence of an open-source framework specialized for financial tasks is imminent.

### 2.4 AI4Finance Foundation and Open-Source Culture

The AI4Finance Foundation<sup>1</sup> plays a crucial role in shaping the open-source culture within financial technology, driven by its mission to promote standardized practices and develop

---

<sup>1</sup><https://www.ai4finance.org>

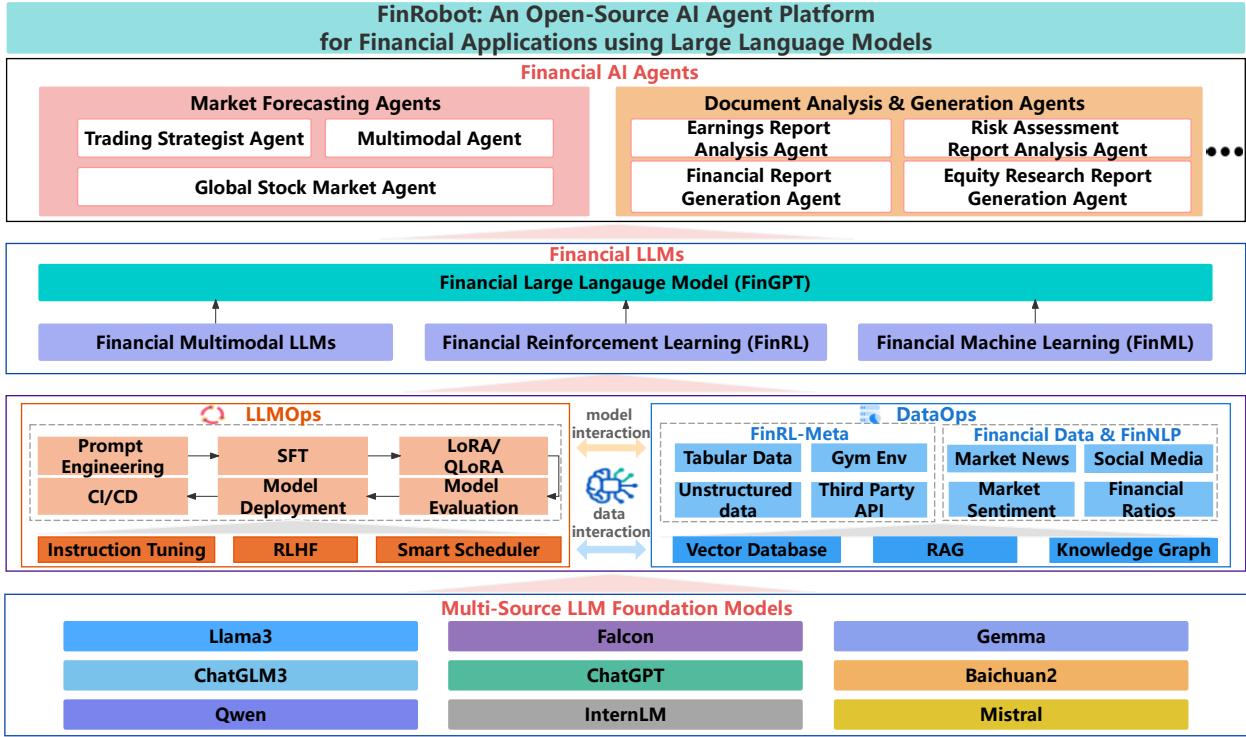


Figure 1: Overall Framework of FinRobot.

open-source resources. This commitment is aimed at benefiting both researchers and industry professionals, fostering a collaborative environment where industry-specific knowledge meets and innovation thrives. By leading initiatives in applying AI technologies for financial services, AI4Finance not only accelerates technological advancements but also ensures these developments are accessible, transparent, and beneficial across the financial sector.

### 3 Overview of FinRobot: An Open-Source Platform for Financial Tasks

As delineated in Fig. 1, the overall framework of FinRobot is organized into four distinct layers, each designed to address specific aspects of financial AI processing and application:

- **Financial AI Agents Layer:** The Financial AI Agents Layer now includes Financial Chain-of-Thought (CoT) prompting, enhancing complex analysis and decision-making capacity. Market Forecasting Agents, Document Analysis Agents, and Trading Strategies Agents utilize CoT to dissect financial challenges into logical steps, aligning their advanced algorithms and domain expertise with the evolving dynamics of financial markets for precise, actionable insights.
- **Financial LLMs Algorithms Layer:** The Financial LLMs Algorithms Layer configures and utilizes specially tuned models tailored to specific domains and global market analysis. It employs FinGPT [Wang *et al.*, 2023d] alongside multi-source LLMs such as

Llama series [Meta, 2023] for the U.S. market and ChatGLM [Zeng *et al.*, 2022] for the Chinese market, each optimized for regional specificities. The Falcon [Almazrouei *et al.*, 2023] model excels in financial relationship analysis. Additionally, multimodal models integrate text with candlestick charts, while FinRL [Yang *et al.*, 2020; Liu *et al.*, 2020] optimizes tasks like portfolio allocation, and traditional machine learning methods [Yang *et al.*, 2018] refine stock selection. This approach ensures high precision in sensitive operations such as market forecasting and financial document analysis.

- **LLMOps and DataOps Layers:** The LLMOps layer implements a multi-source integration strategy that selects the most suitable LLMs for specific financial tasks, utilizing a range of state-of-the-art models. Initially, general LLMs are deployed; if performance is suboptimal, the system dynamically switches to fine-tune LLMs to enhance effectiveness. This adaptive approach ensures tailored solutions for diverse financial scenarios, enhancing the platform's overall performance. Concurrently, the DataOps layer manages real-time data processing [Liu *et al.*, 2022; Liu *et al.*, 2024; Liu *et al.*, 2023], which is crucial for rapid market responsiveness. This dual-layer configuration bolsters FinRobot's ability to deliver timely and accurate financial insights under dynamic market conditions.
- **Multi-source LLM Foundation Models Layer:** This foundational layer supports the plug-and-play function-

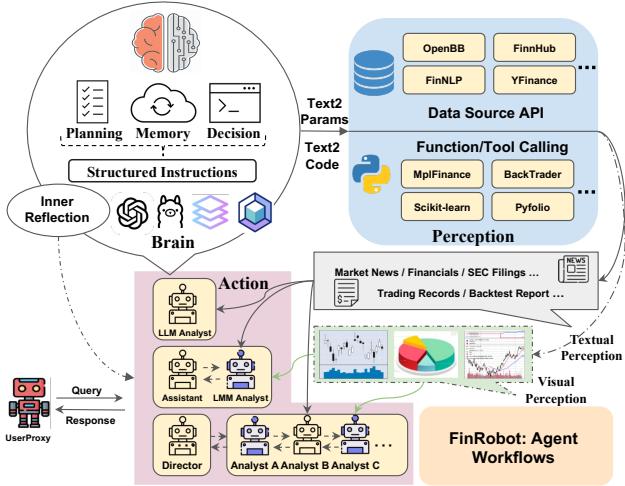


Figure 2: Financial AI Agents Workflow Diagram

ability of various general and specialized LLMs. It forms the platform’s backbone, ensuring that all models are up-to-date, optimized, and consistently aligned with the latest advancements in financial technologies and data standards.

### 3.1 Financial AI Agents Layer

The Financial AI Agents Layer in FinRobot, as depicted in Fig. 2, consists of domain-specific AI agents tailored to enhance financial analysis through advanced data perception, cognitive processing, and dynamic action execution:

**Perception** This module captures and interprets multi-modal financial data from market feeds, news, and economic indicators, using sophisticated techniques to structure the data for thorough analysis.

**Brain** Acting as the core processing unit, this module perceives data from the Perception module with LLMs and utilizes Financial Chain-of-Thought (CoT) processes, as explained in Section 4, to generate structured instructions.

**Action** This module executes instructions from the Brain module, applying tools to translate analytical insights into actionable outcomes. Actions include trading, portfolio adjustments, generating reports, or sending alerts, thereby actively influencing the financial environment.

### Multi-Agent Workflow

Multi-agent workflow systems allow multiple agents, each with unique roles and responsibilities, to tackle intricate tasks in a cooperative manner. The deployment of a multi-agent workflow system is beneficial when navigating complex financial datasets and ensuring a high standard of analytical accuracy and depth. This system incorporates multiple specialized roles that work in concert to process, analyze, and derive actionable insights from diverse financial data sources.

**Director** As the strategic leader of the project, the Director in a financial analysis context oversees all aspects of the

workflow. This role entails prioritizing financial tasks, allocating resources, and orchestrating team efforts to optimize analysis timelines and outcomes.

**Assistant** The Assistant in financial analysis focuses on initial data management tasks such as collecting, processing, and conducting preliminary analyses of financial data and facilitates communication between agents. This role supports more specialized analysts by preparing datasets, maintaining databases, and executing basic financial computations and visualizations. This foundational work enables other analysts to focus on more complex and interpretive aspects of financial analysis.

**LLM Analyst** The LLM Analyst employs advanced computational techniques to analyze financial texts, such as reports, filings, and news articles, using LLMs. This role is essential for extracting detailed insights, conducting sentiment analysis, and forecasting market trends from qualitative data. By integrating these insights with quantitative data, the LLM Analyst provides a comprehensive assessment of financial health and market conditions, enhancing the decision-making process.

**Financial Analysts** Operating under the LLM Analyst’s guidance, Financial Analysts engage in detailed quantitative data analysis across various sectors like portfolio management, risk assessment, and market analysis. Utilizing statistical tools and financial models, they interpret data, evaluate investment opportunities, and formulate risk mitigation strategies. Their expert analysis contributes significantly to the development of sound financial strategies and informed recommendations.

The collaboration among these roles within the multi-agent workflow enables a holistic approach to financial analysis, addressing macroeconomic conditions and specific financial metrics. This structured yet flexible approach ensures thorough exploration and interpretation of financial data, leading to informed strategic decision-making.

### Utilizing LLMs for Tool Usage

**API Interaction through Text2Params** This method translates natural language queries into API requests [Wang *et al.*, 2024], effectively combining the capabilities of generating function calls and forming API calls. In this approach, the LLM first parses the text to identify and extract key parameters. These parameters are then utilized to dynamically generate function calls or compile directly into API requests that interact with financial software or databases. This method is especially useful for executing predefined operations and fetching or manipulating financial data from external services.

**Code Compilation through Text2Code** For more complex financial tasks that require dynamic solution generation, LLMs can employ text2code techniques to write and compile code on the fly. This capability is essential for developing custom algorithms based on user queries’ unique market conditions or financial scenarios.

### 3.2 Financial LLMs Algorithms Layer:

This layer encompasses advanced AI algorithms specifically designed to address various needs within the financial sector, enhancing the platform's capabilities across a range of financial applications.

#### Financial Large Language Models (FinGPT)

FinGPT [Yang *et al.*, 2023], a domain-specific LLM, is meticulously engineered to elevate natural language understanding specifically within the financial context. These models adeptly analyze and interpret financial narratives, extracting critical data from complex documents such as annual reports and real-time financial news, thereby supporting enhanced decision-making processes.

The FinGPT models are mainly developed by supervised finetuning with the financial in-domain paired “instruction-response” data on the open-source large language models. This process is achieved by minimizing the following negative log-likelihood,

$$\mathcal{L}_{\text{CausalLM}} = - \sum_{t=1}^T \log P(w_t | w_1, w_2, \dots, w_{t-1}; \theta), \quad (1)$$

where  $T$  is the length of the input sequence,  $w_i$  represents the  $i$ -th token in the sequence,  $\theta$  denotes the model parameters, and  $P(w_t | w_1, w_2, \dots, w_{t-1}; \theta)$  represents the conditional probability of predicting the target token  $x_i$  given the preceding tokens  $x_1, \dots, x_{t-1}$ . By optimizing this objective function, the model learns to maximize the probability of generating the expected response given the finance task instruction.

#### Financial Reinforcement Learning (FinRL)

FinRL [Yang *et al.*, 2020] optimizes trading strategies using ensemble deep RL algorithms to analyze historical and real-time market data. This dynamic adaptation helps maximize financial returns while minimizing risks, making stock portfolio allocation an ideal application for this methodology.

In FinRL, stock portfolio allocation is modeled as a Markov Decision Process (MDP). At any time  $t$ , an agent in state  $s_t \in \mathcal{S}$  selects an action  $a_t \in \mathcal{A}$  based on policy  $\pi_\theta(s_t)$ . This action leads to a new state  $s_{t+1}$  and a reward  $r(s_t, a_t, s_{t+1})$ . The goal is to optimize this policy:

$$\begin{aligned} \pi_\theta^* &= \underset{\theta}{\operatorname{argmax}} J(\pi_\theta), \\ \text{where } J(\pi_\theta) &= \mathbb{E} \left[ \sum_{t=0}^T \gamma^t r(s_t, a_t, s_{t+1}) \right], \end{aligned} \quad (2)$$

where  $\gamma \in (0, 1]$  is the discount factor.

#### Financial Machine Learning (FinML)

FinML [Yang *et al.*, 2018] leverages diverse machine learning techniques, from regression to advanced neural networks, to boost predictive analytics in finance. These algorithms play a crucial role in forecasting market trends, consumer behavior, credit risks, and other pivotal financial indicators, thus facilitating informed decision-making. A key measure used by FinML is the log-return, calculated as follows:

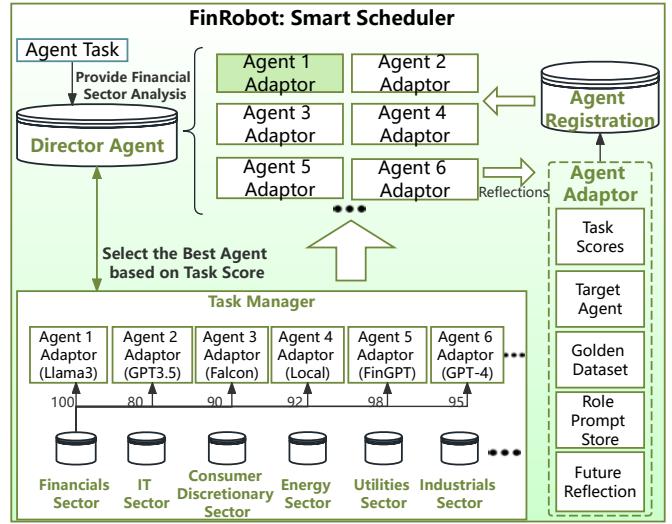


Figure 3: Smart Scheduler Structural Layout in LLMOps Layer

$$r_{T+f,i} = \log \left( \frac{S_{T+f,i}}{S_{T,i}} \right), \quad i = 1, \dots, n_T, \quad (3)$$

where  $r$  denotes the log-return,  $S$  is the stock price,  $n_T$  represents the number of companies monitored at time  $T$ , and  $f$  is the forecast horizon. This measure is essential for evaluating investment performance and shaping financial strategies.

#### Financial Multimodal LLMs

Financial documents frequently incorporate diverse data types beyond textual content, such as graphs [Bhatia *et al.*, 2024] and tables [Wang *et al.*, 2023a]. These additional modalities offer rich, complementary insights, significantly enhancing the depth of analysis possible. To effectively integrate and leverage these varied data types, we have developed Financial Multimodal LLMs, which is specifically engineered to process and synthesize information from multiple modalities, thereby providing a comprehensive and nuanced understanding of financial documents.

The mathematical representation of the integration process in our Financial Multimodal LLM is as follows:

$$F(x_t, x_g, x_h) = L(T(x_t), G(x_g), H(x_h)) \quad (4)$$

where  $F(x_t, x_g, x_h)$  denotes the output of the model. Here,  $x_t$ ,  $x_g$  and  $x_h$  represent the inputs of textual, graphical, and tabular data, respectively. The functions  $T$ ,  $G$  and  $H$  transform these inputs into a unified embedding space. The LLM  $L$  then synthesizes these embeddings to produce a coherent and reliable output, enhancing both the accuracy and reliability of the financial analyses.

### 3.3 LLMOps Layer

The LLMOps Layer is designed for high modularity and pluggability, accommodating rapid integration and dynamic swapping of LLMs in response to evolving technological advancements and financial market demands. This layer facilitates seamless model integration and includes mechanisms for rigorous evaluation and selection of the most suitable

models for specific financial tasks. These capabilities are critical for maintaining operational efficiency and ensuring the adaptability of solutions to diverse financial scenarios.

### Smart Scheduler

The Smart Scheduler, depicted in Figure 3, is central to ensuring model diversity and optimizing the integration and selection of the most appropriate LLM for each task.

### Smart Scheduler Architecture

The Smart Scheduler optimizes task distribution among agents and comprises:

- **Director Agent:** This component orchestrates the task assignment process, ensuring that tasks are allocated to agents based on their performance metrics and suitability for specific tasks.
- **Agent Registration:** Manages the registration and tracks the availability of agents within the system, facilitating an efficient task allocation process.
- **Agent Adaptor:** Tailors agent functionalities to specific tasks, enhancing their performance and integration within the overall system.
- **Task Manager:** Manages and stores different general and fine-tuned LLMs-based agents tailored for various financial tasks, updated periodically to ensure relevance and efficacy.

### Smart Scheduler Initialization Process

The initialization process is streamlined to establish a robust foundation for agent operations effectively:

- **Golden Dataset Creation:** Populates with industry-specific data
- **Prompt Store Population:** Fills with custom prompts for various agents.
- **Task Score Filling:** Invokes adaptors to evaluate responses against best answers, storing scores in the Task Scores database.

### Smart Scheduler In Action Process

The operational phase follows this sequence:

1. **Task Initiation:** A user initiates a task.
2. **Director Agent's Role:** Evaluates task inputs, ranking agents by performance and relevance.
3. **Agent Selection and Task Routing:** Routes tasks to the highest-ranked agent
4. **Workflow Progression and Self-Evaluation:** Post-task completion, the agent performs a self-assessment, the results of which are stored in the Future Reflection storage.
5. **Workflow Completion and Evaluation:** At the end of the workflow, the agent evaluates the results and provides feedback for continuous improvement.

### Scoring Metrics

The scoring process within the Smart Scheduler involves several critical steps to evaluate the performance of different LLMs for task-specific applications:

1. **Data Collection:** Collect performance data of various LLMs across multiple evaluation tasks.
2. **Normalization:** Normalize the results for each evaluation task to scale between 0 and 1.
3. **Weight Assignment:** Assign weights to different evaluation dimensions based on industry standards or expert opinions.
4. **Calculation of Composite Score:** Multiply the normalized scores by their respective weights and sum them to derive the task score.
5. **Results Analysis:** Analyze the scores to rank and assess the LLMs, providing performance comparisons and recommendations for selection.

This enhanced structure not only boosts the operational efficiency of financial AI agents but also supports scalable, dynamic management of diverse financial tasks, significantly improving the system's capability to manage complex, multi-agent scenarios in real-time financial environments.

### 3.4 DataOps Layer

The DataOps Layer manages the extensive and varied datasets necessary for financial analysis, ranging from public tabular data to proprietary market sentiments. This layer is crucial in ensuring that all data fed into the AI processing pipelines is of high quality and representative of the current market conditions. Effective data management is foundational for the accurate and reliable performance of AI models, and as such, the DataOps Layer employs advanced data handling and processing techniques to prepare and deliver these inputs. By optimizing data accessibility and quality, the DataOps Layer supports the overall efficacy of the FinRobot platform, enabling sophisticated and precise financial decision-making based on solid data-driven insights.

### Retrieval Augmented Generation

Retrieval-augmented generation (RAG) is a pivotal technique incorporated within FinGPT [Zhang *et al.*, 2023b; Zhang *et al.*, 2023a], as it seamlessly combines the strength of both context retrieval mechanisms and LLMs to optimize language generation tasks.

The LLMOps and DataOps layers form the backbone of FinRobot's operational capabilities, ensuring that the platform remains at the forefront of technological innovation while providing reliable and effective financial AI services.

### 3.5 Multi-source LLM Foundation Models Layer

The Multi-source LLM Foundation Models Layer equips FinRobot with advanced capabilities to manage and integrate a diverse array of Large Language Models (LLMs), essential for adapting to the dynamic demands of global financial markets. Key features of this layer include:

- **Plug-and-Play Functionality:** Enables seamless integration and updating of both general and specialized LLMs, ensuring the platform remains adaptable and up-to-date with financial technology advancements.
- **Model Diversity and Evaluation:** Incorporates LLMs with parameters ranging from 7 billion to 72 billion, each rigorously evaluated for effectiveness in specific financial tasks, allowing for optimal model selection based on performance metrics such as accuracy and adaptability.
- **Global Market Compatibility:** Supports multilingual model integration, enhancing the platform's ability to analyze and process diverse financial data, which is crucial for global market operations.

This streamlined version succinctly outlines the layer's functionality and strategic role in maintaining FinRobot's leadership in AI-driven financial analysis.

## 4 Financial Chain-of-Thought (CoT) Prompting

### 4.1 Introduction

The chain-of-thought prompting technique [Wei *et al.*, 2022; Xia *et al.*, 2024] structures prompts to foster a step-by-step reasoning process within AI models, akin to human problem-solving strategies. This method significantly enhances performance on complex reasoning tasks such as mathematics and commonsense reasoning by encouraging models to articulate intermediate reasoning steps, leading to a final answer. This method improves accuracy and enhances the interpretability and transparency of the decision-making process.

### 4.2 Concept of Financial CoT

Financial CoT Prompting adapts the CoT technique for AI-driven financial analysis, integrating advanced cognitive processing techniques to refine decision-making capabilities. It stimulates the thought process of financial professionals to tackle complex financial issues by guiding AI models through a logical, sequential reasoning process. This approach breaks down complex financial scenarios into smaller, manageable components, analyzes each component, and synthesizes the findings to form conclusions or recommendations. As noted by [Kim *et al.*, 2024], this method mimics human-like reasoning in the analysis chain and is particularly valuable for tasks requiring deep analytical thought, such as valuation, investment strategy formulation, market trend analysis, and risk assessment.

### 4.3 Financial CoT Implementation

Financial Chain-of-Thought (CoT) Prompting marks a transformative shift in security analysis, enhancing traditional analytical tools across various domains by leveraging the capabilities of LLMs such as GPT-4. Detailed case studies, such as those conducted by [Kim *et al.*, 2024], highlight this approach's practical applications and benefits.

- **Financial Analysis:** FinRobot transcends basic data enumeration and ratio computation. By employing

LLMs for financial statement analysis, it conducts comparative analyses across industry competitors and against the company's historical performance. This comprehensive approach provides profound insights into a company's business landscape and the logical underpinnings of its financial metrics. Specifically, it normalizes financial ratios to identify anomalies, enhancing the accuracy of financial insights beyond human capabilities.

- **Business-Specific Analysis:** Leveraging Retrieval-Augmented Generation (RAG), FinRobot revolutionizes the process of gathering data on a company's products and services, complementing the information available in financial reports. By tapping into the vast resources of the web, FinRobot enriches its understanding of a company's offerings, enabling a comprehensive analysis of product lines, trends across channels and regions, cost structures, supply chain dynamics, and R&D conversions.
- **Market Analysis:** FinRobot abstracts real-world financial contexts, utilizing a combination of financial ratios, market data (including stock price trends and candlestick charts), market news sentiment, and alternative data sourced online. By employing LLMs, FinRobot simulates market participants' decision-making processes in response to contextual changes. This allows for a comprehensive evaluation of a company's stock price trajectory and valuation. FinRobot provides nuanced investment assessments, offering strategic guidance on optimal investment timing and suitable financial instruments, ensuring robust, well-informed investment strategies considering various horizons, structures, and risk thresholds.
- **Valuation Analysis:** FinRobot integrates its comprehensive analysis of financial ratios, market data (including stock price trends and candlestick charts), market news sentiment, and alternative data collected online. FinRobot employs LLMs to evaluate a company's stock price trajectory and valuation using the insights derived from these diverse sources. This approach enables FinRobot to provide nuanced investment assessments across various horizons, structures, and risk thresholds, offering strategic guidance on optimal investment timing and suitable financial instruments, thereby ensuring well-informed and robust investment strategies.

### 4.4 Advantages of Financial CoT Prompting

Despite their merits, a readily available tool that can intelligently perform industry- and company-specific analytics still needs to exist in a way that parallels the work of an investment professional. Existing data providers such as Bloomberg, FactSet, CapitalIQ, and Refinitiv Eikon rely on rudimentary, generic templates for data extraction, inadvertently overlooking key operational metrics and qualitative information intricately linked with the business essence. Manual analysis can be time-consuming and occasionally repetitive while existing automated tools can only sometimes extract raw data from

unstructured sources like annual reports. Although these platforms abound with vast amounts of information, extracting essential information and its subsequent analysis still require the expertise of investment professionals.

With CoT Prompting, FinRobot establishes a unique niche among AI Agent Platforms by emulating human cognitive processes in financial analysis, transcending traditional reliance on mere numerical computations. Our methodology leverages LLMs to deconstruct and analyze the methodologies utilized by financial professionals when scrutinizing business data. This approach unveils a practical analysis intrinsically dependent on a comprehensive understanding of a company's business fundamentals, as reflected through detailed accounting data, tailored industry-specific metrics, and qualitative information.

This approach overcomes the above limitations because: (1) It does not rely on a rigid template due to the generative nature of LLMs. (2) It avoids misclassification because error-checking is embedded within layers of prompting. (3) It simplifies human labor by codifying investment logic within its reasoning. (4) It can extract relevant historical numbers and qualitative information from complex and unstructured data formats such as earnings call transcripts. (5) Given its multi-layered structure, it provides stronger grounds for explaining the sources and derivations of each recorded and derived value. (6) It is highly adaptable and evolving as LLMs may be improved through supervised fine-tuning and instruction-tuning.

#### 4.5 Market Simulation: Transcending the Analyst Mimicking

To expand the concept of market simulation and explore its possibilities, one could view the GPT's ability to mimic human analytical processes as a gateway to simulating broader market dynamics. Market simulation transcends mere numerical analysis by incorporating human-like reasoning processes, embodying multiple personas and scenarios within a financial context. This simulation could encompass a variety of market participants whose decisions are influenced by an evolving set of data inputs—ranging from financial statements, reports, market data, and economic indicators to global news.

The core of this approach involves constructing a detailed simulation environment where these generative agents, representing different market players, interact [Park *et al.*, 2023]. Each agent's decision-making process would be guided by data-driven prompts, mimicking the cognitive steps a human analyst might take. This includes interpreting complex market signals and reacting to new information in real time, which advanced machine learning models and reinforcement learning techniques could facilitate.

By leveraging LLMs, one could theoretically encode these decision-making frameworks into the agents, enabling them to perform tasks that require a deep understanding of the market's quantitative and qualitative aspects. This would simulate individual trading decisions and the broader market dynamics resulting from multiple decision-makers interactions. The potential to use this technology to create a virtual market environment where hypothetical scenarios could be tested

and analyzed could provide invaluable insights into market behavior and strategy development.

A detailed framework for such a simulation could involve: (1) Defining the roles and behaviors of different market participants (e.g., institutional investors, retail investors, market makers). (2) Implementing a variety of data inputs and decision-making models for each participant. (3) Simulating market interactions over different time horizons and under varying conditions to assess potential outcomes and strategies. (4) Utilizing reinforcement learning to refine the agents' strategies based on observed performance in the simulation.

As outlined by [Park *et al.*, 2023], this concept presents a promising avenue for future research and development in financial market analysis and strategy optimization.

## 5 Demo Applications and Hands-on Tutorial

### 5.1 Application I: Market Forecaster

Market Forecaster is a set of AI Agents designed to synthesize recent market news and financial data, delivering comprehensive insights into a company's latest achievements and potential concerns, along with predictions for stock price movements. As a junior robo-advisor, Market Forecaster embodies a substantial leap forward in AI-driven financial advisory.

**Instruction:** You are an experienced stock market analyst. Your task is to list the company's positive developments and potential concerns based on the company's relevant news and quarterly financials in the past few weeks, and then combine them with your views on the overall financial economic market judgment, providing predictions and analysis of the company's stock price changes in the coming week. Your answer format should be as follows:

[Positive development]:

1. ...

[Potential concerns]:

1. ...

[Forecast and Analysis]:

...

**Information:**

- a. Company Introduction
- b. Stock Price Changes
- c. Recent News Information
- d. Recent Basic Financials

**Instruction:** Based on all the information before 2024-04-19, let's first analyze the positive developments and potential concerns for AAPL. Come up with 2-4 most important factors respectively and keep them concise. Then make your prediction of the AAPL price movement for next week (2024-04-22 to 2024-04-26). Provide a summary analysis to support your prediction.

Figure 4: Market Forecaster Prompt Template

### Data

Market Forecaster stands at the crossroads of global markets, leveraging diverse data sources to gather information

### Forecaster for NVDA on 2024-01-29

#### [Positive Developments]:

1. NVDA's stock price has been steadily increasing over the past weeks, suggesting a strong investor sentiment. ([Stock Price](#))
2. The company's CEO, Jensen Huang, has expressed optimism about the future of AI and Nvidia's role in it, which could boost investor confidence. ([News](#))
3. The company has received attention from Zacks.com users, indicating a strong performance in the market. ([News](#))
4. The company has a strong balance sheet with a healthy gross margin and a high EV. ([Basic Financials](#))

#### [Potential Concerns]:

1. The company's long-term debt-to-total asset and debt-to-total capital ratios are relatively high, indicating a potential risk of financial instability. ([Basic Financials](#))
2. The company's high PE ratio and high net debt-to-total equity ratio suggest high valuation and potential financial risk. ([Basic Financials](#))
3. The company's quick ratio is relatively low, indicating a potential liquidity risk. ([Basic Financials](#))

#### [Prediction & Analysis]:

Prediction: Up by 0-1%

Analysis: Despite some potential concerns, the positive developments for NVDA are significant. The company's strong balance sheet and increasing stock price suggest a strong investor sentiment. The optimistic outlook from the CEO and the attention from Zacks.com users also suggest a positive future outlook. However, the high long-term debt-to-total asset and debt-to-total capital ratios are a cause for concern. These high ratios could potentially increase the company's financial risk and affect its liquidity.

Figure 5: Market Forecaster Result of Nvidia

and make forecasting decisions. In detail, Market Forecaster gathers data regarding multifaceted company information such as recent news, latest basic financials, and stock prices for the target markets, including the US stock market, Chinese stock market, Crypto market, and other possible extensions. This section will give a solid overview of the Market Forecaster targeted at the US stock market and Chinese stock market with data sources Finnhub and EastMoney.

#### Model

This AI Agent is powered by fine-tuned FinGPT-Forecasters using Llama-2-7b-chat-hf with LoRA, leveraging data from the latest year's DOW 30 for the US Market and SSE 50 for the Chinese Market to ensure precise forecasts for these major stocks and demonstrating robust generalization abilities across various stock symbols. Moreover, this paper follows the Financial Multi-task Instruction Tuning paradigm [Wang *et al.*, 2023d; Yu *et al.*, 2023b] to align the power of the base model with the specific functions of Market Forecaster.

#### Prompts

Aligning with the multi-task instruction tuning framework, the Market Forecaster follows a sophisticated prompt format. Specifically, the Market Forecaster gathers data of multifaceted company information, then it performs prompt en-

### Forecaster for Kweichow Moutai on 2024-01-28

#### [积极发展]:

1. 贵州茅台的购买率持续上涨，表明市场对茅台的需求持续增长。([News](#))
2. 公司的新闻发布，如“美茅台美世界”2024全球品牌文化活动，表明公司继续追求国际化发展，并且将继续以“美”的国际表达，坚持追求美，实现看好未来的发展。([News](#))
3. 茅台的营业总收入同比增长率达到了19.80%，表明公司的经营状况良好，具有稳定的盈利能力。([Basic Financials](#))
4. 茅台的流动比率和速动比率都相对较高，表明公司的资金流动性较好，可以更好地应对市场的压力。([Basic Financials](#))

#### [潜在担忧]:

1. 茅台的资产负债率达到了17.98%，显示公司的负债压力较大，可能会对公司的盈利能力产生影响。([Basic Financials](#))
2. 公司未来可能面临的市场竞争压力可能会增加，如果公司无法继续提供高质量的产品和服务，可能会受到市场反应。([News](#))
3. 茅台的购买率上涨，可能会导致市场价格的偏高，可能会影响到公司的股价。([Stock Price](#))

#### [预测和分析]:

预测涨跌幅: 上涨0-1%

总结分析: 虽然贵州茅台的资产负债率较高，但公司的经营状况良好，营业收入同比增长率达到了19.80%，显示公司在市场的需求和竞争中具有稳定的盈利能力。考虑到茅台的购买率持续上涨，表明市场对茅台的需求持续增长，可能会对公司的股价产生积极影响。同时，公司的新闻发布，如“美茅台美世界”2024全球品牌文化活动，表明公司继续追求国际化发展，可能会带来未来的发展机会。因此，预计茅台的股价在下一周将上涨0-1%。

Figure 6: Market Forecaster Result of Kweichow Moutai

gineering to format the instructive prompt with the structure of ‘Task Instruction & Company Information (Company Overview + Recent Stock Prices + Recent News + Latest Basic Financials)’. Figure 4 shows an example of the Market Forecaster prompt template.

#### Sample Forecastings

The sample answers provided for two prominent stocks in the US and Chinese markets, namely Nvidia and Kweichow Moutai, as depicted in Figures 5 and 6, respectively, illustrate the Market Forecaster’s adeptness in synthesizing and offering valuable insights gleaned from diverse information sources. Additionally, the Market Forecaster furnishes recommendations regarding the future trajectory of the stocks, underscoring its capability to provide actionable guidance based on the analyzed data.

## 5.2 Application II: Document Analysis & Generation

The Document Analysis & Generation application represents a groundbreaking use of AI Agents and Large Language Models (LLMs) in the realm of financial document management and report creation. This application harnesses the power of AI to perform deep analysis of financial documents and generate detailed, insightful reports automatically.

## Document Analysis

AI Agents, integrated with advanced LLMs, are employed to sift through extensive financial documents such as annual reports, SEC filings, and earnings call transcripts. These agents are capable of extracting critical information, identifying key financial indicators, and highlighting trends and discrepancies that may require closer inspection. The ability of these AI Agents to understand and process complex financial terminology and context turns vast amounts of unstructured data into structured, actionable insights.

## Report Generation

Following the analysis phase, the same agent is utilized to generate comprehensive financial reports. Leveraging the capabilities of LLMs, the system produces coherent, articulate, and detailed documents that cover various aspects of financial analysis, including performance evaluations, market comparisons, and forward-looking financial forecasts. Each report is crafted to maintain a professional tone and format, mirroring the quality and depth expected of top-tier financial analysts.

For practical insights, equity research reports generated by FinRobot are included in the appendix, showcasing the application and effectiveness of our platform.

## 6 Conclusion

FinRobot revolutionizes financial analysis by integrating multi-source Large Language Models (LLMs) in an open-source platform that enhances accessibility, efficiency, and transparency in financial operations. This innovative platform addresses the complexities of global markets with a multi-layered architecture that supports real-time data processing and diverse model integration, making sophisticated financial tools available to a wider audience. By fostering collaboration within the financial AI community, FinRobot not only accelerates innovation but also sets new standards for the application of AI in finance, promising to significantly improve strategic decision-making across the sector.

**Future Work** As FinRobot continues to evolve, we plan to expand its applications to include more sophisticated tasks such as portfolio allocation and comprehensive risk assessment, enhancing its utility in the financial sector. Additionally, we aim to broaden the platform's reach by extending its capabilities to more global markets. This expansion will not only diversify the potential use cases of FinRobot but also improve its adaptability to different economic environments, driving innovation and accessibility in AI-driven financial analysis across the world.

## References

- [Abarbanell and Bushee, 1997] Jeffrey S Abarbanell and Brian J Bushee. Fundamental analysis, future earnings, and stock prices. *Journal of accounting research*, 35(1):1–24, 1997.
- [Almazrouei *et al.*, 2023] Ebtesam Almazrouei, Hamza Alobeidli, Abdulaziz Alshamsi, Alessandro Cappelli, Ruxandra Cojocaru, Merouane Debbah, Etienne Goffinet, Daniel Heslow, Julien Launay, Quentin Malartic, Badreddine Noune, Baptiste Pannier, and Guilherme Penedo. Falcon-40B: an open large language model with state-of-the-art performance. *arXiv*, 2023.
- [Awotunde *et al.*, 2021] Joseph Bamidele Awotunde, Emmanuel Abidemi Adeniyi, Roseline Oluwaseun Ogun-dokun, and Femi Emmanuel Ayo. Application of big data with fintech in financial services. In *Fintech with artificial intelligence, big data, and blockchain*, pages 107–132. Springer, 2021.
- [Baek *et al.*, 2024] Jinheon Baek, Sujay Kumar Jauhar, Silviu Cucerzan, and Sung Ju Hwang. Researchagent: Iterative research idea generation over scientific literature with large language models. *arXiv preprint arXiv:2404.07738*, 2024.
- [Berman and Knight, 2013] Karen Berman and Joe Knight. *Financial intelligence, revised edition: A manager’s guide to knowing what the numbers really mean*. Harvard Business Review Press, 2013.
- [Bhatia *et al.*, 2024] Gagan Bhatia, El Moatez Billah Nagoudi, Hasan Cavusoglu, and Muhammad Abdul-Mageed. Fintral: A family of gpt-4 level multimodal financial large language models. *arXiv preprint arXiv:2402.10986*, 2024.
- [Brown *et al.*, 2020] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901, 2020.
- [Cao, 2022] Longbing Cao. Ai in finance: challenges, techniques, and opportunities. *ACM Computing Surveys (CSUR)*, 55(3):1–38, 2022.
- [Durante *et al.*, 2024] Zane Durante, Qiuyuan Huang, Naoki Wake, Ran Gong, Jae Sung Park, Bidipta Sarkar, Rohan Taori, Yusuke Noda, Demetri Terzopoulos, Yejin Choi, et al. Agent ai: Surveying the horizons of multimodal interaction. *arXiv preprint arXiv:2401.03568*, 2024.
- [Gao *et al.*, 2023] Chen Gao, Xiaochong Lan, Nian Li, Yuan Yuan, Jingtao Ding, Zhilun Zhou, Fengli Xu, and Yong Li. Large language models empowered agent-based modeling and simulation: A survey and perspectives. *arXiv preprint arXiv:2312.11970*, 2023.
- [Gravitas, 2023] Significant Gravitas. Autogpt. <https://github.com/Significant-Gravitas/AutoGPT>, 2023.
- [Greenwald *et al.*, 2004] B.C. Greenwald, J. Kahn, P.D. Sonkin, and M. van Biema. *Value Investing: From Graham to Buffett and Beyond*. Business book summary. Wiley, 2004.
- [Henrique *et al.*, 2019] Bruno Miranda Henrique, Vini-cius Amorim Sobreiro, and Herbert Kimura. Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, 124:226–251, 2019.
- [Hong *et al.*, 2023] Sirui Hong, Mingchen Zhuge, Jonathan Chen, Xiawu Zheng, Yuheng Cheng, Ceyao Zhang, Jinlin Wang, Zili Wang, Steven Ka Shing Yau, Zijuan Lin,

- [Liyang Zhou, Chenyu Ran, Lingfeng Xiao, Chenglin Wu, and Jürgen Schmidhuber. Metagpt: Meta programming for a multi-agent collaborative framework, 2023.]
- [Huang *et al.*, 2023] Allen H Huang, Hui Wang, and Yi Yang. Finbert: A large language model for extracting information from financial text. *Contemporary Accounting Research*, 40(2):806–841, 2023.
- [Jiang, 2021] Weiwei Jiang. Applications of deep learning in stock market prediction: recent progress. *Expert Systems with Applications*, 184:115537, 2021.
- [Kim *et al.*, 2024] Alex Kim, Maximilian Muhn, and Valeri V Nikolaev. Financial statement analysis with large language models. *Chicago Booth Research Paper Forthcoming, Fama-Miller Working Paper*, 2024.
- [Kumar *et al.*, 2022] Deepak Kumar, Pradeepa Kumar Sarangi, and Rajit Verma. A systematic review of stock market prediction using machine learning and statistical techniques. *Materials Today: Proceedings*, 49:3187–3191, 2022.
- [LangGenius, 2024] LangGenius. Dify. <https://github.com/langgenius/dify>, 2024.
- [Liu *et al.*, 2020] Xiao-Yang Liu, Hongyang Yang, Qian Chen, Runjia Zhang, Liuqing Yang, Bowen Xiao, and Christina Dan Wang. FinRL: A deep reinforcement learning library for automated stock trading in quantitative finance. *Deep RL Workshop, NeurIPS 2020*, 2020.
- [Liu *et al.*, 2022] Xiao-Yang Liu, Ziyi Xia, Jingyang Rui, Jiechao Gao, Hongyang Yang, Ming Zhu, Christina Dan Wang, Zhaoran Wang, and Jian Guo. FinRL-Meta: Market environments and benchmarks for data-driven financial reinforcement learning. *NeurIPS*, 2022.
- [Liu *et al.*, 2023] Xiao-Yang Liu, Guoxuan Wang, Hongyang Yang, and Daochen Zha. Data-centric fingpt: Democratizing internet-scale data for financial large language models. *NeurIPS Workshop on Instruction Tuning and Instruction Following*, 2023.
- [Liu *et al.*, 2024] Xiao-Yang Liu, Ziyi Xia, Hongyang Yang, Jiechao Gao, Daochen Zha, Ming Zhu, Christina Dan Wang, Zhaoran Wang, and Jian Guo. Dynamic datasets and market environments for financial reinforcement learning. *Machine Learning - Nature*, 2024.
- [Medhat *et al.*, 2014] Walaa Medhat, Ahmed Hassan, and Hoda Korashy. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4):1093–1113, 2014.
- [Meta, 2023] Meta. LLaMA 2: Open foundation and fine-tuned chat models. *Preprint*, 2023.
- [Murphy, 1999] John J Murphy. *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*. Penguin, 1999.
- [Nabipour *et al.*, 2020] Mojtaba Nabipour, Pooyan Nayyeri, Hamed Jabani, Amir Mosavi, Ely Salwana, and Shahab S. Deep learning for stock market prediction. *Entropy*, 22(8):840, 2020.
- [Park *et al.*, 2023] Joon Sung Park, Joseph O’Brien, Carrie Jun Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pages 1–22, 2023.
- [Penman, 2010] S. Penman. *Accounting for Value*. Columbia Business School Publishing. Columbia University Press, 2010.
- [Qian *et al.*, 2023] Chen Qian, Xin Cong, Wei Liu, Cheng Yang, Weize Chen, Yusheng Su, Yufan Dang, Jiahao Li, Juyuan Xu, Dahai Li, Zhiyuan Liu, and Maosong Sun. Communicative agents for software development, 2023.
- [Qiao *et al.*, 2023] Bo Qiao, Liquan Li, Xu Zhang, Shilin He, Yu Kang, Chaoyun Zhang, Fangkai Yang, Hang Dong, Jue Zhang, Lu Wang, et al. Taskweaver: A code-first agent framework. *arXiv preprint arXiv:2311.17541*, 2023.
- [Rasheed *et al.*, 2024] Zeeshan Rasheed, Muhammad Waseem, Mika Saari, Kari Systä, and Pekka Abrahamsson. Codepori: Large scale model for autonomous software development by using multi-agents. *arXiv preprint arXiv:2402.01411*, 2024.
- [Shen *et al.*, 2024] Yongliang Shen, Kaitao Song, Xu Tan, Dongsheng Li, Weiming Lu, and Yueling Zhuang. Hugginggpt: Solving ai tasks with chatgpt and its friends in hugging face. *Advances in Neural Information Processing Systems*, 36, 2024.
- [Sohangir *et al.*, 2018] Sahar Sohangir, Dingding Wang, Anna Pomeranets, and Taghi M Khoshgoftaar. Big data: Deep learning for financial sentiment analysis. *Journal of Big Data*, 5(1):1–25, 2018.
- [Sumers *et al.*, 2023] Theodore R Sumers, Shunyu Yao, Karthik Narasimhan, and Thomas L Griffiths. Cognitive architectures for language agents. *arXiv preprint arXiv:2309.02427*, 2023.
- [Tang *et al.*, 2023] Xiangru Tang, Anni Zou, Zhuosheng Zhang, Yilun Zhao, Xingyao Zhang, Arman Cohan, and Mark Gerstein. Medagents: Large language models as collaborators for zero-shot medical reasoning. *arXiv preprint arXiv:2311.10537*, 2023.
- [Thirunavukarasu *et al.*, 2023] Arun Thirunavukarasu, Darren Shu Jeng Ting, Kabilan Elangovan, Laura Gutierrez, Ting Fang Tan, and Daniel Shu Wei Ting. Large language models in medicine. *Nature medicine*, 29(8):1930–1940, 2023.
- [Wang *et al.*, 2023a] Dongsheng Wang, Natraj Raman, Mathieu Sibue, Zhiqiang Ma, Petr Babkin, Simerjot Kaur, Yulong Pei, Armeneh Nourbakhsh, and Xiaomo Liu. Doclmm: A layout-aware generative language model for multimodal document understanding. *arXiv preprint arXiv:2401.00908*, 2023.
- [Wang *et al.*, 2023b] Guanzhi Wang, Yuqi Xie, Yunfan Jiang, Ajay Mandlekar, Chaowei Xiao, Yuke Zhu, Linxi Fan, and Anima Anandkumar. Voyager: An open-ended embodied agent with large language models. *arXiv preprint arXiv:2305.16291*, 2023.

- [Wang *et al.*, 2023c] Lei Wang, Chen Ma, Xueyang Feng, Zeyu Zhang, Hao Yang, Jingsen Zhang, Zhiyuan Chen, Jiakai Tang, Xu Chen, Yankai Lin, et al. A survey on large language model based autonomous agents. *arXiv preprint arXiv:2308.11432*, 2023.
- [Wang *et al.*, 2023d] Neng Wang, Hongyang Yang, and Christina Dan Wang. Fingpt: Instruction tuning benchmark for open-source large language models in financial datasets. *NeurIPS Workshop on Instruction Tuning and Instruction Following*, 2023.
- [Wang *et al.*, 2024] Zhiruo Wang, Zhoujun Cheng, Hao Zhu, Daniel Fried, and Graham Neubig. What are tools anyway? a survey from the language model perspective. *arXiv preprint arXiv:2403.15452*, 2024.
- [Wei *et al.*, 2022] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.
- [Wu *et al.*, 2023a] Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, Ahmed Hassan Awadallah, Ryen W White, Doug Burger, and Chi Wang. Autogen: Enabling next-gen llm applications via multi-agent conversation framework. In *arXiv*, 2023.
- [Wu *et al.*, 2023b] Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabrowski, Mark Dredze, Sebastian Gehrmann, Prabhjanan Kambadur, David Rosenberg, and Gideon Mann. BloombergGPT: A large language model for finance. *arXiv preprint arXiv:2303.17564*, 2023.
- [Xi *et al.*, 2023] Zhiheng Xi, Wenxiang Chen, Xin Guo, Wei He, Yiwen Ding, Boyang Hong, Ming Zhang, Junzhe Wang, Senjie Jin, Enyu Zhou, et al. The rise and potential of large language model based agents: A survey. *arXiv preprint arXiv:2309.07864*, 2023.
- [Xia *et al.*, 2024] Yu Xia, Rui Wang, Xu Liu, Mingyan Li, Tong Yu, Xiang Chen, Julian McAuley, and Shuai Li. Beyond chain-of-thought: A survey of chain-of-x paradigms for llms. *arXiv preprint arXiv:2404.15676*, 2024.
- [Yang *et al.*, 2018] Hongyang Yang, Xiao-Yang Liu, and Qingwei Wu. A practical machine learning approach for dynamic stock recommendation. *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/ 12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, pages 1693–1697, 2018.
- [Yang *et al.*, 2020] Hongyang Yang, Xiao-Yang Liu, Shan Zhong, and Anwar Walid. Deep reinforcement learning for automated stock trading: An ensemble strategy. In *Proceedings of the first ACM international conference on AI in finance*, pages 1–8, 2020.
- [Yang *et al.*, 2023] Hongyang Yang, Xiao-Yang Liu, and Christina Dan Wang. Fingpt: Open-source financial large language models. *FinLLM Symposium at IJCAI 2023*, 2023.
- [Yu *et al.*, 2023a] Dingyao Yu, Kaitao Song, Peiling Lu, Tianyu He, Xu Tan, Wei Ye, Shikun Zhang, and Jiang Bian. Musicagent: An ai agent for music understanding and generation with large language models. *arXiv preprint arXiv:2310.11954*, 2023.
- [Yu *et al.*, 2023b] Xinli Yu, Zheng Chen, and Yanbin Lu. Harnessing llms for temporal data-a study on explainable financial time series forecasting. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages 739–753, 2023.
- [Yu *et al.*, 2023c] Yangyang Yu, Haohang Li, Zhi Chen, Yuechen Jiang, Yang Li, Denghui Zhang, Rong Liu, Jordan W. Suchow, and Khaldoun Khashanah. Finmem: A performance-enhanced llm trading agent with layered memory and character design, 2023.
- [Zeng *et al.*, 2022] Aohan Zeng, Xiao Liu, Zhengxiao Du, Zihan Wang, Hanyu Lai, Ming Ding, Zhuoyi Yang, Yifan Xu, Wendi Zheng, Xiao Xia, et al. Glm-130b: An open bilingual pre-trained model. *arXiv preprint arXiv:2210.02414*, 2022.
- [Zeng *et al.*, 2023] Fanlong Zeng, Wensheng Gan, Yongheng Wang, Ning Liu, and Philip S Yu. Large language models for robotics: A survey. *arXiv preprint arXiv:2311.07226*, 2023.
- [Zhang *et al.*, 2023a] Boyu Zhang, Hongyang Yang, and Xiao-Yang Liu. Instruct-fingpt: Financial sentiment analysis by instruction tuning of general-purpose large language models. *FinLLM Symposium at IJCAI 2023*, 2023.
- [Zhang *et al.*, 2023b] Boyu Zhang, Hongyang Yang, Tianyu Zhou, Ali Babar, and Xiao-Yang Liu. Enhancing financial sentiment analysis via retrieval augmented large language models. *ACM International Conference on AI in Finance (ICAIF)*, 2023.
- [Zhang *et al.*, 2024] Wentao Zhang, Lingxuan Zhao, Haochong Xia, Shuo Sun, Jiaze Sun, Molei Qin, Xinyi Li, Yuqing Zhao, Yilei Zhao, Xinyu Cai, et al. Finagent: A multimodal foundation agent for financial trading: Tool-augmented, diversified, and generalist. *arXiv preprint arXiv:2402.18485*, 2024.

**Disclaimer:** We are sharing codes for academic purposes under the MIT education license. They should not be construed as financial counsel or recommendations for live trading. It is imperative to exercise caution and consult with qualified financial professionals prior to any trading or investment actions.

# Equity Research Report: NVIDIA Corporation

## Income Summarization

In fiscal year 2024, NVIDIA witnessed a remarkable revenue increase to \$60.922 billion, up 126% from the previous year, driven predominantly by a 217% surge in Data Center revenue, highlighting the robust demand for its computing platforms. The Compute & Networking segment, in particular, saw revenue skyrocket to \$47.405 billion, up 215%, with significant contributions from the NVIDIA Hopper GPU computing platform for AI applications. Meanwhile, the Graphics segment also experienced growth, with a 14% increase in revenue to \$13.517 billion, bolstered by a 15% rise in Gaming revenue due to the launch of new GPUs and enhanced AI-powered gaming experiences. The company's gross profit margin improved to 72.7%, and net profit margin expanded to 48.9%, reflecting cost efficiency, a favorable product mix, and operational effectiveness. Despite challenges such as increased operational costs and supply chain complexities, NVIDIA's strategic partnerships and continuous innovation in AI and gaming technologies position it for sustained growth and market leadership.

## Business Highlights

- Data Center: Revenue surged by 217% due to strong demand for NVIDIA Hopper GPU computing platforms for AI applications. This growth was driven by the increasing adoption of AI across various industries, including healthcare and automotive.
- Gaming: Revenue increased by 15%, driven by higher sell-in to partners following the normalization of channel inventory levels. The growth reflects the continued demand for gaming GPUs and the introduction of new products.
- Professional Visualization: Revenue saw a modest increase of 1%, indicating stable demand in the professional graphics market. This stability is attributed to ongoing demand for NVIDIA's visualization products across industries.
- Automotive: Revenue grew by 21%, primarily due to growth in self-driving platforms. This growth signifies NVIDIA's expanding presence in the automotive sector, particularly in AI and autonomous driving technologies.

## Company Situation

NVIDIA Corporation, a leader in accelerated computing, thrives on innovation in GPU technology and AI, catering to diverse sectors including AI, data analytics, and 3D graphics. Its competitive edge is bolstered by significant R&D investments, leading to breakthroughs like the GPU and deep learning algorithms. Current trends like the demand for AI, the rise of the metaverse, and the need for high-performance computing present both opportunities and challenges, including supply chain issues and intense competition. NVIDIA's strategic initiatives, such as launching new GPU architectures and expanding into data center CPUs, alongside partnerships in autonomous driving, aim to sustain its industry leadership by continuously innovating and adapting to market needs.

## Risk Assessment

The top 3 key risks for NVIDIA Corporation include: 1) Failure to adapt to rapid changes in technology and customer requirements, which could adversely impact financial results; 2) Intense competition that could reduce market share and financial performance; and 3) Regulatory, legal, and political actions, including export controls and tariffs, that may restrict business operations or increase operational costs, particularly those affecting the semiconductor and AI industries, which could materially impact the company's ability to manufacture, sell, and distribute products globally.

## Financial Metrics

FY (USD mn)	2020	2021	2022	2023	2024
Operating Revenue	10918.0	16675.0	26914.0	26974.0	60922.0
Adjusted Net Profit	2796.0	4332.0	9752.0	4368.0	29760.0
Adjusted EPS	1.15	1.76	3.91	1.76	12.05
EBIT Margin	0.26	0.27	0.37	0.16	0.54
ROE	0.23	0.26	0.37	0.2	0.69
PE Ratio	52.32	75.41	62.67	109.1	51.82
EV/EBITDA	40.53	58.59	54.7	66.08	44.84
PB Ratio	11.99	19.34	22.97	21.56	35.88

## FinRobot

<https://ai4finance.org/>

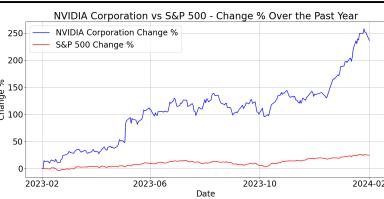
<https://github.com/AI4Finance-Foundation/FinRobot>

Report date: 2024-02-21

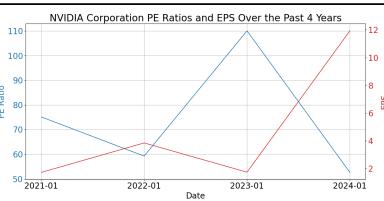
## Key data

Rating	hold
Target Price	850 - 1100 (md. 875.0)
6m avg daily val (USDmn)	46.27
Closing Price (USD)	694.49
Market Cap (USDmn)	1663859.52
52 Week Price Range (USD)	210.94 - 724.98
BVPS (USD)	17.41

## Share Performance



## PE & EPS



## Revenue

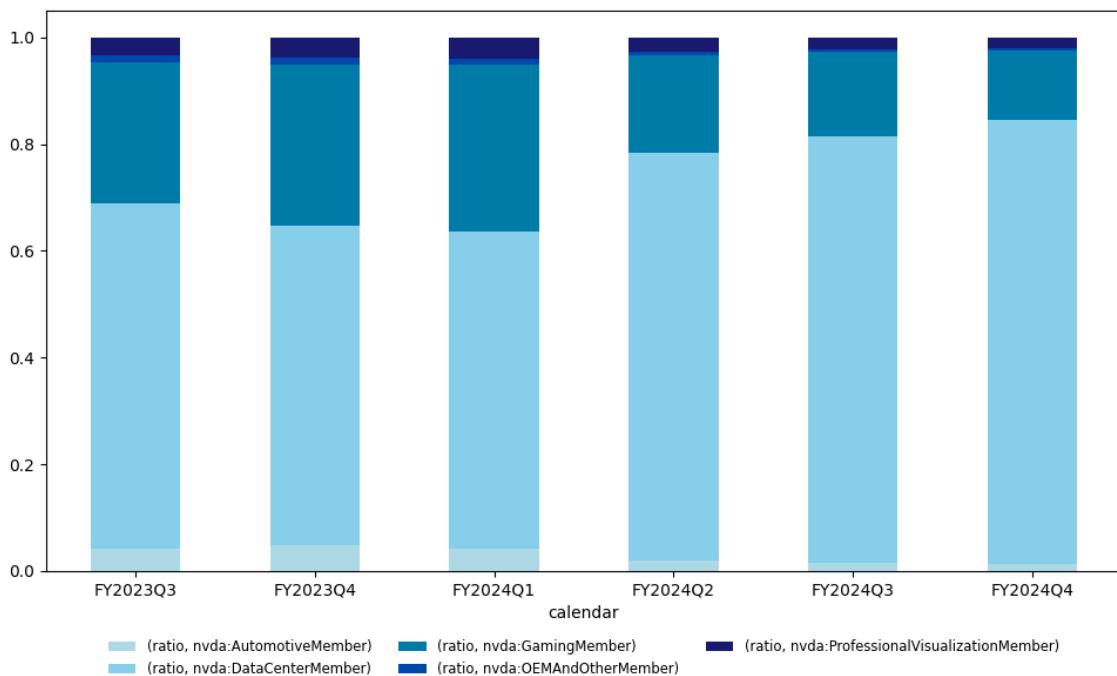
segment	FY2023Q3	FY2023Q4	FY2024Q1	FY2024Q2	FY2024Q3	FY2024Q4
AutomotiveMember	251.00	294.00	296.00	253.00	261.00	281.00
DataCenterMember	3833.00	3616.00	4284.00	10323.00	14514.00	18404.00
GamingMember	1574.00	1831.00	2240.00	2486.00	2856.00	2865.00
OEMAndOtherMember	73.00	84.00	77.00	66.00	73.00	90.00
ProfessionalVisualizationMember	200.00	226.00	295.00	379.00	416.00	463.00
total	5931.00	6051.00	7192.00	13507.00	18120.00	22103.00

## Ratio

segment	FY2023Q3	FY2023Q4	FY2024Q1	FY2024Q2	FY2024Q3	FY2024Q4
AutomotiveMember	0.04	0.05	0.04	0.02	0.01	0.01
DataCenterMember	0.65	0.60	0.60	0.76	0.80	0.83
GamingMember	0.27	0.30	0.31	0.18	0.16	0.13
OEMAndOtherMember	0.01	0.01	0.01	0.00	0.00	0.00
ProfessionalVisualizationMember	0.03	0.04	0.04	0.03	0.02	0.02

## YoY

segment	FY2023Q3	FY2023Q4	FY2024Q1	FY2024Q2	FY2024Q3	FY2024Q4
AutomotiveMember	86.00	135.00	114.00	15.00	4.00	-4.00
DataCenterMember	31.00	11.00	14.00	171.00	279.00	409.00
GamingMember	-51.00	-46.00	-38.00	22.00	81.00	56.00
OEMAndOtherMember	-69.00	-56.00	-51.00	-53.00	0.00	7.00
ProfessionalVisualizationMember	-65.00	-65.00	-53.00	-24.00	108.00	105.00
total	-17.00	-21.00	-13.00	101.00	206.00	265.00



# 证券研究报告：贵州茅台

## 核心观点

2023年，公司实现营收1476.94亿元，同比增长19.01%，归母净利润747.34亿元，增长19.16%。这一强劲的业绩表明，尽管市场环境充满挑战，公司依然能够通过优化销售策略和调整产品矩阵，持续增长和盈利。特别是，公司对直销渠道的重视促进了营收的显著增加，而在产品方面，系列酒的快速增长开始改变收入结构，为公司的持续发展注入新动力。这些成绩反映了公司不仅在行业中保持了领先地位，同时也拥有良好的市场竞争力。对于未来，公司通过进一步优化直销渠道和扩展产品矩阵的策略，展现出对增强市场竞争力和持续增长能力的明确信心。

## 业绩总览

公司发布了2023年的主要经营数据：营收达到约1476.94亿元，同比增长约19.01%；归属于母公司的净利润为747.34亿元，同比增长约19.16%。该数据显示公司在面临挑战的市场环境中继续展现出强劲的成长动力和盈利能力，稳健的经营业绩体现了其在行业中的领先地位和良好的市场竞争力。

## 销售模式

根据提供的数据，公司的销售渠道包含直销和批发代理两大部分。从2019年至2023年数据分析，直销营收呈现显著增长趋势，增速从65.65%逐年攀升至高峰的105.5%，尽管在2023年增速回落至36.16%，但直销营收仍然保持强劲增长。相比之下，批发代理营收增长相对稳定，除2022年出现-9.31%的下滑外，其他年份增速较低。此数据显示公司销售策略逐年偏向直销渠道，直销渠道对收入的贡献日益增大，而批发代理渠道的影响相对减弱。直销渠道的强劲增长可能是因为直接与消费者的互动提升了销售效率和客户满意度，也反映了市场需求和消费者偏好的变化。

## 产品矩阵

根据所提供数据，公司产品矩阵围绕“茅台酒”和“系列酒”。茅台酒作为核心产品，其收入持续占据主导地位，2019-2023年收入增速呈上升趋势，由15.75%增至17.39%。系列酒虽然起点较低，但增速更快，尤其从2020年开始，其增速显著跃升，2023年达到29.43%。这显示系列酒在产品矩阵中的比重正在增长，对收入的贡献越来越显著。因此，虽然茅台酒仍是收入的主要来源，但系列酒的快速增长正在逐步改变公司的收入结构，显示公司正通过扩展产品矩阵增加收入的多样性和稳定性。

## 金融指标

报告期	2019	2020	2021	2022	2023
营业收入	888.54亿	979.93亿	1094.64亿	1275.54亿	1505.60亿
营业总收入同比增长率	15.10%	10.29%	11.71%	16.53%	18.04%
净利润	412.06亿	466.97亿	524.60亿	627.17亿	747.34亿
净利润同比增长率	17.05%	13.33%	12.34%	19.55%	19.16%
销售净利率	51.47%	52.18%	52.47%	52.68%	52.49%
销售毛利率	91.30%	91.41%	91.54%	91.87%	91.96%
基本每股收益	32.8000	37.1700	41.7600	49.9300	59.4900
净资产收益率	33.09%	31.41%	29.90%	30.26%	34.19%
市盈率	42.3138	58.9284	55.8193	41.3543	34.5714
市净率	11.8922	16.357	14.9506	12.1073	11.1874

## FinRobot

<https://ai4finance.org/>

<https://github.com/AI4Finance-Foundation/FinRobo>

Report date: 2024-04-03

## 基本数据

评级	买入
目标价	1953.65 - 2600.0 (md. 2246.9)
近六月日平均成交量 (RMBmn)	0.03
收盘价 (RMB)	1715.11
总市值 (RMBbn)	2125499.24
52周价格范围 (RMB)	1555.55 - 1935.00

## 股价表现



## 市盈率和每股收益

