LLM-Powered Multi-Agent System for Automated Crypto Portfolio Management

Anonymous ACL submission

Abstract

Cryptocurrency investment is non-trivial due to its short history, the involvement of multi-modal data, and the need for complex reasoning. While deep learning has addressed some of these challenges, its "black-box" nature limits trust and explainability. Recently, large language models (LLMs) have shown promise in financial applications by effectively understanding multi-modal data and generating explainable decisions. However, single LLM faces limitations in complex, comprehensive tasks such as asset investment. These limitations are even more pronounced in cryptocurrency investment, where LLMs often lack sufficient domain-specific knowledge within their training corpora.

To overcome these challenges, we propose an explainable, multi-modal, multi-agent framework for cryptocurrency investment. Our framework uses specialized agents that collaborate within and across teams to handle subtasks such as data analysis, literature integration, and investment decision-making for the top 30 cryptocurrencies by market capitalization using multi-modal data. Unique intrateam and interteam collaboration mechanisms enhance predictability by adjusting final predictions based on confidence levels within agent teams and facilitating information sharing between teams. Empirical evaluation using data from November 2023 to September 2024 demonstrates that our framework outperforms single-agent models and market benchmarks in multiple metrics. Our framework delivers an annualized cumulative return of 108.32% and an annualized Sharpe ratio of 1.5425 in portfolio performance and generates statistically significant long-short portfolio returns in asset pricing.

1 Introduction

Cryptocurrency investment is a challenging and comprehensive task due to its limited asset pricing evidence (Corbet et al., 2019; Fang et al., 2022), the requirement for data from various modalities (Jing and Kang, 2024; Kapur et al., 2024; Liu and Tsyvinski, 2021; Liu et al., 2022), and the need for complex reasoning (Hackethal et al., 2022). As a result, analyzing the cryptocurrency market, designing strategies, and building portfolios become a huge undertaking and impose a heavy workload on financial experts, making professional services either scarce or expensive (CFP Board, 2022). To address these challenges, many researchers have explored the use of deep learning techniques (Goutte et al., 2023; Lahmiri and Bekiros, 2019) for cryptocurrency investment. However, the "black-box" nature of most deep learning models raises concerns about trust and explainability, making investors hesitant to rely on these techniques when investing their capital (Li et al., 2023b; Carta et al., 2021; Biran and McKeown, 2017).

The introduction of LLMs has revolutionized the financial field, offering promising solutions for cryptocurrency investment. Numerous studies have demonstrated the strong capability of LLMs to understand and learn from multi-modal data (Yin et al., 2024) such as text (Yang et al., 2024; Li et al., 2023c) and images (Zhang et al., 2024), which makes them well-suited for learning professional cryptocurrency investment knowledge and analyzing the market from data in different modalities. On the other hand, LLMs has excellent natural language generation capability (Wei et al., 2022; Liu et al., 2023a), which enables them to generate explainable cryptocurrency investment decisions. However, the performance of single LLMs in asset prediction is limited due to the comprehensive nature and complex reasoning requirement of this task (Xie et al., 2023; Li et al., 2023a). The weakness is even more pronounced in cryptocurrency investment, where LLMs have less domain-specific knowledge in their training corpora. To address this type of challenge, researchers have developed methodologies that decompose complex tasks into subtasks (Wu et al., 2023; Pan et al., 2024). This approach uses the collaboration between multiple LLM-based agents to derive final comprehensive solutions, with each agent focusing on a specific aspect of the overall task. Inspired by human cognitive processes, this approach enhances reasoning capabilities and efficacy in solving comprehensive problems, offering new possibilities for agent-based investment solutions. Although some studies have explored the use of multi-agent models in stock investment (Ding et al., 2024; Fatemi and Hu, 2024; Oprea and Bara, 2025), works that employ the multi-agent model in cryptocurrency investment are few and are limited to Bitcoin, Ethereum, and Solana as well as data in single modality (Li et al., 2024; Yu et al., 2025).

To address the above-mentioned problems and fill in the gap, we propose an explainable, multi-modal, multi-agent framework, which utilizes multiple teams of agents that collaborate both within and across teams to facilitate supervised learning and investment decisions across the top 30 cryptocurrencies by market capitalization. Within this framework, complex investment tasks involving data from different modalities are decomposed into several subtasks, with each fine-tuned expert agent assigned responsibility for a specific subtask. Inspired by the communication methods used in hedge funds, our unique intrateam and interteam collaboration mechanism ensures that the final investment decision integrates information from multiple modalities effectively.

Our multi-agent framework consists of two modules: the expert training module (Figure 1a) and the multi-agent investment module (Figure 1c). The expert training module employs agents from the data team and literature team to fetch historical multi-modal data and relevant investment literature, respectively. Next, agents in the explanation team process the data and literature to generate high-quality prompts by integrating multi-modal information and professional investment knowledge. Finally, these prompts are used to fine-tune expert investment agents, each specializing in the analysis of data in a single modality. The multi-agent investment module utilizes the data team to fetch real-time data and forward it to the market team and crypto team. The market team includes two expert agents who analyze news and market factors to predict market trends and determine the cash-crypto allocation. Similarly, the crypto team has two expert agents who analyze crypto-specific factors and candlestick charts of individual cryptocurrencies to make crypto selection decisions. Finally, the trading team interacts with cryptocurrency exchange APIs to execute the final portfolio strategy. The intrateam collaboration mechanism combines the confidence scores of agents within the same group to produce an ensemble prediction. The interteam collaboration mechanism allows agents in the crypto team to share memory with the market team regarding market information, enabling more robust crypto selection decisions based on comprehensive information.

To demonstrate the effectiveness of our framework, we use data from June 2023 to September 2024 to validate its ability to outperform single-agent models, both with and without fine-tuning, in terms of asset pricing and explainability. Additionally, we show that our framework surpasses market benchmarks in portfolio performance. The main contributions of this paper are summarized as:

- We are the first to employ an LLM-powered multi-agent framework for managing large-cap cryptocurrency portfolios. Our framework not only outperforms single LLMs in asset pricing performance but also surpasses market benchmarks in portfolio performance.

- We design a unique multi-agent framework that enables both interteam and intrateam communication among agents. These mechanisms facilitate communication and mitigate prediction errors among different agents, significantly enhancing the performance of our model in cryptocurrency investment.

- Our multi-agent framework integrates multi-modal financial data such as text and vision data into decision-making. This design enables the model to achieve a holistic understanding of the market dynamics, improving prediction accuracy and investment decisions.

- We design unique asset pricing methods for LLM to convert the binary rise-or-fall price trend classification into a spectrum of confidence levels using the token probability. These confidence levels are then used to build portfolios according to the empirical asset pricing methodology in finance.

2 Related Works

In this section, we examine works that utilize single LLM and multi-agent frameworks for investment.

With their powerful text understanding and reasoning capabilities, LLMs have become widely used in different investment tasks. Early studies have focused on employing single LLMs to predict asset prices and execute investment strategies. Some works have attempted to fine-tune their own financial LLM to complete investment tasks (Xie et al., 2023; Liu et al., 2023b; Li et al., 2023a; Oprea and Bara, 2025). Additionally, some studies specifically examined the performance of LLMs in trading three cryptocurrencies: Bitcoin, Ethereum, and Solana (Li et al., 2024; Yu et al., 2025). However, the predictive power of single LLMs remains limited even after fine-tuning, and their results often exhibit significant bias.

To further improve the performance of LLM in investment, recent research has shifted towards using multi-agent models for investment tasks. One notable example is the Summarize-Explain-Predict (SEP) framework, which employs a reflective agent that iteratively generates stock predictions and explanations with assistance from other agents (Koa et al., 2024). Some studies focus on using multiple agents to process data, summarize information, reflect, and generate stock prediction, respectively (Kou et al., 2024; Fatemi and Hu, 2024; Ding et al., 2024). However, there remains a gap in the development of multi-agent, multi-modal models specifically designed for cryptocurrency investment tasks. To fill in this gap, we propose a multi-agent framework where specialized agents, each responsible for processing distinct modalities of information, collaboratively invest in a universe of leading cryptocurrencies.

3 Methodology

In this section, we first decompose the cryptocurrency investment process into multiple subtasks and formalize them. Next, we introduce a multi-agent cryptocurrency investment framework designed to address these subtasks collaboratively. Finally, we compare our multi-agent system with conventional single-agent systems.

3.1 Problem Formulation

3.1.1 Cryptocurrency-cash allocation

We first use information from news and market data to predict market booms and busts, and allocate cash and cryptocurrency to minimize the impact of extreme market events (i.e., systematic risk). Given a vector of market-specific risk factors $\boldsymbol{\beta}_{t-1} = [\beta_{i,t-1}]_{p\times 1}$ at week $t-1$, where p denotes the total number of factors, and news data $\mathbf{N}_{t-1} = [N_{i,t-1}]_{q\times 1}$ at week $t-1$, where q denotes the total number of news headlines, our goal is to generate the crypto weight w_t to maximize the weighted market return: $\underset{w_t}{\operatorname{argmax}}\, w_t r_t^{\text{mkt}}$ and a human-readable explanation $\hat{e}_t^{\text{mkt}}$ at week t.

3.1.2 Cryptocurrency selection

We then employ the crypto-specific information to predict the crypto price trends (i.e., unsystematic risk). Given a set of cryptocurrencies c in $\mathcal{C} = \{c_i\}_{i=1}^{I}$, a matrix of crypto-specific risk factors $\boldsymbol{\alpha}_{c_i,t-1} = [\alpha_{j,c_i,t-1}]_{m\times n}$, where m is the total number of crypto-specific risk factors and n is the total number of cryptos, and a vector of visual data $\mathbf{v}_{c_i,t-1} = [v_{c_i,t-1}]_{m\times 1}$, we aim to generate a subset $\mathcal{C}^* \subseteq \mathcal{C}$ to maximize the average future 7-day returns of those cryptos $\underset{\mathcal{C}^*\subseteq\mathcal{C}}{\operatorname{argmax}}\, \frac{1}{|C^*|}\sum_{c_i\in C^*} r_t^{c_i}$ and a human-readable explanation $\hat{e}_t^{c_i}$.

3.1.3 Portfolio weighting

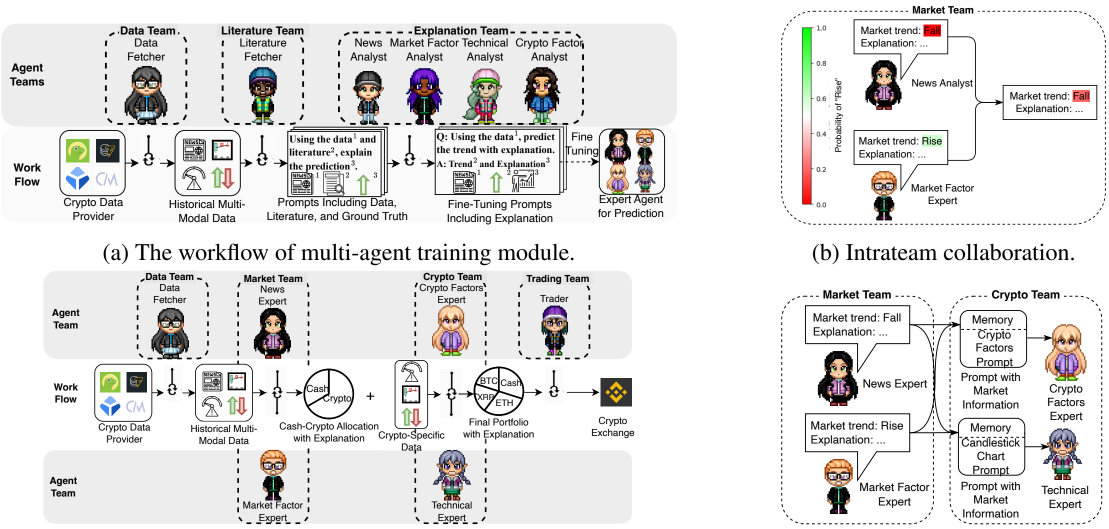
We finally use the crypto-specific information mentioned in the cryptocurrency selection task to decide the weights w_{c_i} to maximize the future 7-day returns of the selected subset of cryptos $\underset{\mathcal{C}^*\subseteq\mathcal{C}}{\operatorname{argmax}}\, w_{c_i}\sum_{c_i\in C^*} r_t^{c_i}$.

3.2 Framework Overview

In this paper, we propose an explainable multi-agent framework for cryptocurrency investment, as illustrated in Figure 1. Our framework consists of two major modules: multi-agent training and multi-agent investment. As shown in Figure 1a, the multi-agent training module generates training prompts that incorporate data from various modalities along with corresponding high-quality, case-by-case explanations. Subsequently, knowledge derived from diverse data modalities is integrated into the respective expert agents through fine-tuning. As illustrated in Figure 1c, the mult-agent investment module enables expert agents to manage corresponding subtasks in the cryptocurrency investment and collaboratively construct the final cryptocurrency portfolio.

3.3 Expert Training

The expert training process involves collaboration among multiple agent teams. Within the data team, the data fetcher is responsible for fetching and processing raw data. This agent utilizes tools to gather



(a) The workflow of multi-agent training module. (b) Intrateam collaboration.

(c) The workflow of multi-agent investment module. (d) Interteam collaboration.

Figure 1: Multi-agent framework for automated cryptocurrency portfolio management.

data from leading cryptocurrency providers, including Coingecko, Blockchain.info, Coin Metrics, and Cointelegraph. Once the raw data is fetched, the data fetcher processes it into multi-modal formats, including price trend ground truth, 30-day candlestick charts, risk factors (alphas), and news headlines, as illustrated by Table 4 and Figure 5 in Appendix A. The 30-day candlestick charts (Figure 5a) and binary price trends (Figure 5b) are derived from open high low and close (OHLC) price data and trading volume provided by Coingecko. Risk factors (Figure 5c) are computed using OHLC price data, trading volume, and market capitalization from Coingecko, along with on-chain data sourced from Blockchain.info and Coin Metrics. Risk factors are categorized into five quintiles: "Very Low", "Low", "Medium", "High", and "Very High", as described in Equation 4. The quintile cutoffs are determined using cross-sectional data for crypto-related factors and the initial two years of data for market-related factors. Additionally, news headline data (Figure 5d) is obtained through web crawling from Cointelegraph.

In the literature team, the literature fetcher is tasked with retrieving academic papers from Google Scholar. The academic papers are selected based on their relevance to specific data modalities in the domain of empirical cryptocurrency pricing. To generate training prompts with detailed case-by-case reasoning derived from academic papers, an explanation team is responsible for enhancing the training data. This team transforms plain training data pairs, consisting of multi-modal data and ground truth, into enriched pairs by incorporating professional, well-reasoned explanations. Specifically, the market factor analyst and news analyst focus on analyzing market-specific risk factors and news data from the current week, along with the corresponding market trend for the following week. Using the shared System Instruction 1 and Prompts 1, 2 in Appendix C, the market factor and news analysts explain the complex relationships between market-related information and the market trend, leveraging insights from relevant academic papers. Similarly, the crypto analyst uses the System Instruction 1 and Prompt 3 to analyze crypto-specific risk factors and ground truth, generating detailed explanations. Then, the technical analyst employs the System Instruction 1 and Prompt 4 to interpret the relationship between 30-day candlestick charts of individual cryptocurrencies and their corresponding ground truth, providing well-reasoned insights. Finally, the multi-modal data, ground truth, and corresponding explanation are integrated into training prompts using the System Instruction 2 and template Prompt 5. Finally, prompts enhanced by four explanation analysts are fed into four LLMs to train experts in the market team and the crypto team.

3.4 Multi-Agent Investment

The multi-agent investment component employs collaboration among multiple agents to complete the cryptocurrency investment process. This process begins with the data team fetching and processing real-time multi-modal data from various providers. Subsequently, the market team, crypto team, and trading team receive the processed data and complete their respective subtasks, contributing to the overall investment process. The agents start with zero crypto plus a given level of cash.

To complete the cryptocurrency-cash allocation subtask, the market team employs a trained news expert and a trained market factor expert agent A^{news} to predict market trends. Specifically, the news expert is provided with the System Instruction in 3 and a prompt generated by filling news headline data from the past week, $\mathbf{N}_{t-1}$, into the template outlined in Prompt 6. Using this prompt, the news expert generates a prediction for the current week, $\hat{Y}_t^{\text{news}}$, which includes two components: 1. a prediction in either "Rise" or "Fall", representing the expected market trend for the upcoming week and 2. a human-readable explanation, $\hat{e}_t^{\text{news}}$, that provides detailed reasoning behind the prediction, i.e., $\hat{Y}_t^{\text{news}} = (\hat{y}_t^{\text{news}}, \hat{e}_t^{\text{news}})$. We can formalize this process as

$$\hat{Y}_t = \mathrm{A}(X_{t-1}^{\text{mkt}}), \tag{1}$$

where X_{t-1}^{mkt} is the generalized market-specific data for the last week. In this scenario, $X_{t-1}^{\text{mkt}} = \mathbf{N}_{t-1}$. Similarly, the market factor expert agent A^{mf} is provided with the system instruction and prompt integrated with market-specific risk factors, $\boldsymbol{\beta}_{t-1}$, to generate a prediction $\hat{Y}_t^{\text{mf}}$ including a prediction in either "Rise" or "Fall" and a human-readable explanation, $\hat{e}_t^{\text{mf}}$.

To enable collaboration between these two agents within a team and generate a final solution for the crypto-cash allocation subtask, the intrateam collaboration method illustrated in Figure 1b is employed. This method allows the two agents to ensemble their predictions based on their respective prediction confidence levels. Specifically, since the LLM generates text by selecting tokens with the highest probabilities, we can extract the log probability of "Rise" for the classification token ("Rise" or "Fall"), which is expressed as $\ln P(\hat{y}_t = \text{“Rise”} | \mathbf{N}_{t-1})$. This probability serves as an additional confidence measure for the prediction. To ensemble the predictions, the log probabilities are converted into linear probabilities, which is visually represented in Figure 1b, where a greener background indicates a higher log probability. The final ensemble rise probability is then calculated by taking the arithmetic mean of the linear probabilities from both agents, effectively combining their insights to produce a more robust prediction for the subtask:

$$\bar{P} = \frac{1}{|\mathcal{A}|} \sum_{i \in \mathcal{A}} e^{\ln P(\hat{y}_t^i = \text{“Rise”} | X_{t-1})}, \tag{2}$$

where $\mathcal{A}$ represents the set of agents within a team. This aggregated probability is subsequently employed as a weighting factor in portfolio construction, determining the capital allocation to the corresponding cryptocurrency. The residual capital is retained in cash, thereby implementing a dynamic crypto–cash allocation strategy.

To complete the cryptocurrency selection subtask, the crypto team employs a trained crypto factors expert and a trained technical expert to collaboratively predict the price trend for individual cryptocurrencies. To enable the agents in the crypto team to make predictions that incorporate not only crypto-specific information but also the broader market context, we employ interteam collaboration, as illustrated in Figure 1d. Specifically, the agents in the crypto team receive the System Instruction 3 and the Prompt 6 in Appendix C integrated with the relevant crypto-specific data as their primary input. Additionally, they are provided with inputs and predictions from the market factor expert and the news expert, which serve as contextual information or shared short-term memory to enhance their decision-making process. We can formalize it as:

$$\hat{Y}_{c,t} = \mathrm{A}\left(c, X_{c,t-1}, X_{t-1}^{\text{mf}}, \hat{Y}_t^{\text{mf}}, X_{t-1}^{\text{news}}, \hat{Y}_t^{\text{news}}\right). \tag{3}$$

While market information alone does not directly contribute to cross-sectional cryptocurrency price trend prediction, since all cryptocurrencies share the same market-level data, we expect the expert agents to learn the interactions between market-level information and individual crypto. By identifying these interactions, the crypto team can enhance the accuracy of their predictions. Therefore, the crypto factors expert is provided with individual crypto c and the vector of its crypto-specific risk factors, $\boldsymbol{\alpha}_{c,t-1} = [\alpha_{i,c,t-1}]_{m \times 1}$, while the technical expert receives the crypto c and its 30-day candlestick chart of, $\mathbf{v}_{c,t-1}$. Using these inputs, the experts generate either "Rise" or "Fall", $\hat{y}_t^{\text{cf}} \in \{\text{“Rise”}, \text{“Fall”}\}$ and $\hat{y}_t^{\text{chart}} \in \{\text{“Rise”}, \text{“Fall”}\}$, representing the predicted price trends for the cryptocurrencies over the following week. Additionally, they produce human-readable explanations, $\hat{e}_t^{\text{cf}}$ and

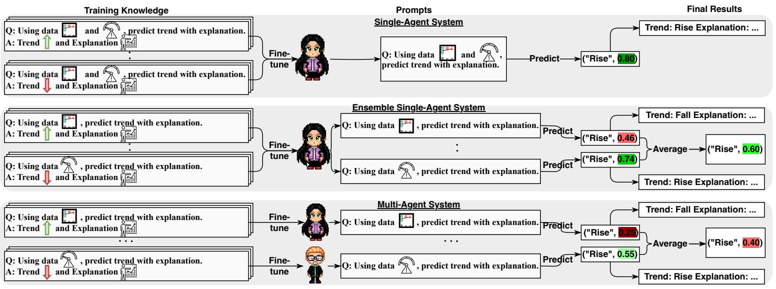


Figure 2: A comparison of the single-agent, ensemble single-agent, and multi-agent system using the example of the market team in Figure 1c.

Table 1: Comparison of portfolio results in full, boom, and bust periods.

Period	Portfolio	Mean	Std	Sharpe
All	Ours	0.0172	0.0805	1.5425
	Market	0.0131	0.0683	1.3781
	1/N	0.0082	0.0834	0.7070
	Bitcoin	0.0144	0.0677	1.5340
Boom	Ours	0.0428	0.0802	3.8430
	Market	0.0422	0.0630	4.8279
	1/N	0.0391	0.0758	3.7195
	Bitcoin	0.0404	0.0634	4.5977
Bust	Ours	-0.0269	0.0715	−2.7125
	Market	-0.0359	0.0529	−4.8944
	1/N	-0.0479	0.0723	−4.7748
	Bitcoin	-0.0299	0.0519	−4.1570

$\hat{e}_t^{\text{chart}}$, providing detailed reasoning behind their respective predictions.

Using the same intrateam collaboration method, expert agents within the crypto group come to a consensus about the price trend of individual cryptocurrencies by generating the final ensemble rise probability for each individual cryptocurrency $c \in \mathcal{C}$, $\bar{P}_c$, via Equation 2. Then, we sort cryptocurrencies in set $\mathcal{C}$ into quintile portfolios based on the $\bar{P}_c$. Specifically, we form 5 disjoin equal-weighted (1/N) portfolios, each representing a range of rise probabilities, denoted by $\mathcal{P}_i$. The portfolios are constructed as follows:

$$\mathcal{P}_i = \left[\bar{P}_{\left(\left\lfloor \frac{|\mathcal{C}|(i-1)}{5} \right\rfloor\right)}, \bar{P}_{\left(\left\lfloor \frac{|\mathcal{C}|i}{5} \right\rfloor\right)} \right) \quad i = 1, \cdots, 5, \tag{4}$$

where $\bar{P}_{(j)}$ denotes the j-th order static of the ascending set of rise probabilities $\{\bar{P}_c : c \in \mathcal{C}\}$ of all cryptocurrencies in set $\mathcal{C}$. $\lfloor\rfloor$ is the floor operator. Portfolios $\mathcal{P}_1, \cdots, \mathcal{P}_5$ are labeled Very Low, Low, Medium, High, and Very High. Finally, the portfolio labeled Very High, $\mathcal{P}_5$, is selected as the target subset of cryptocurrencies, $\mathcal{C}^* = \mathcal{P}_5 \subseteq \mathcal{C}$, for investment.

To complete the portfolio weights subtask, we use $\bar{P}_c$ to construct the probability-weighted portfolio. The final trading team is tasked with executing trades by interacting with the APIs of cryptocurrency exchanges based on the provided portfolio, ensuring that the entire process is fully end-to-end.

3.5 Single Agent versus Multi-Agent System

In this research, we use a multi-agent system to address the limitations of single-agent models, particularly their weakness in weighting diverse sources of information. Figure 2 compares the single-agent and multi-agent approaches. Both the regular and ensemble single-agent systems are fine-tuned on a broad universe of crypto investment knowledge. The key difference lies in their prompting strategies: the former generates predictions for the subtasks using a single prompt for each, while the latter uses multiple granular prompts corresponding to different data types. In contrast, the multi-agent system consists of multiple agents, each fine-tuned with domain-specific knowledge and generating predictions based on corresponding domain-specific prompts. As noted in (Koa et al., 2024), LLM struggles to effectively weigh and integrate different types of information when forming aggregate predictions in financial contexts. This limitation motivates us to empirically evaluate the performance of single-agent versus multi-agent approaches in cryptocurrency investment.

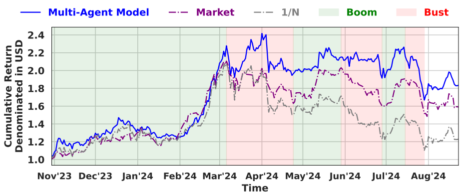
4 Experiment

In this section, we evaluate the performance of our multi-agent framework on our collected dataset against the related baselines.

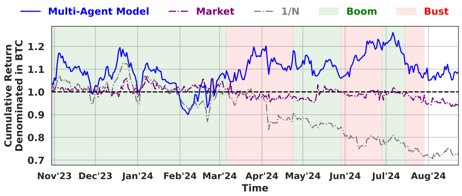
4.1 Experiment Settings

In this work, we employ GPT-4o as the base model, as it is the most advanced multi-modal model capable of implementing vision fine-tuning at the time of writing.[1] We collected our dataset from June 2023 to September 2024. We designate the set of targeted cryptocurrencies, $\mathcal{C}$, as the top 30 cryptocurrencies by market capitalization according to CoinGecko. This list is updated weekly to reflect changes in market capitalization. The rationale for including only high-capitalization cryptocurrencies is that those with low market capitalization often exhibit pricing dynamics that differ significantly from high-liquidity cryptocurrencies, partly due to

[1] https://platform.openai.com/docs/guides/fine-tuning#which-models-can-be-fine-tuned



(a) Denominated in US Dollar.



(b) Denominated in Bitcoin.

Figure 3: Performance comparisons in out-of-sample cumulative returns of our multi-agent model portfolio against baselines.

risks such as pump-and-dump schemes. Additionally, trading low-liquidity cryptocurrencies tend to involve higher slippage, further complicating our task. To prevent information leakage, we set the data from November 2023 to September 2024 as the test set, given that GPT-4o's training data extends only up to October 2023.[2] Consequently, the training set comprises data from June 2023 to October 2023.

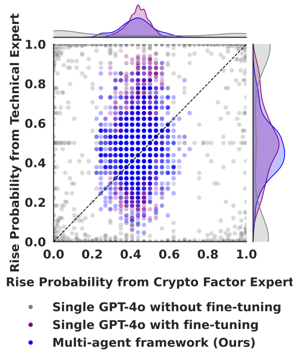
4.2 Performance Comparison

In this subsection, we evaluate the portfolio performance and the asset pricing performance of our multi-agent model via quantitative comparisons against the related baselines.

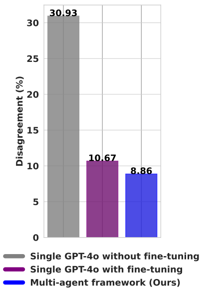
4.2.1 Portfolio Performance

In this section, we evaluate the portfolio performance of our multi-agent framework. Figure 3 depicts the out-of-sample cumulative returns of our multi-agent model against the market index and equal-weighted portfolios (1/N) of the top 30 cryptocurrencies. From Figure 3a, we observe that the portfolio generated by our model outperforms two constructed indices throughout the entire sample period, except for February 2024.

Table 1 presents the comparison of portfolio results across full, boom, and bust periods. The table shows that the portfolio produced by our multi-agent model outperforms other portfolio strategies in most metrics across all periods, including boom and bust phases, while maintaining comparable performance in terms of standard deviation. We provide detailed explanations for metrics used to evaluate portfolios in Appendix D. Notably, our model exhibits strong resistance to declines during the bust period, further highlighting its effectiveness.



(a) Rise probability .



(b) Average disagreement.

Figure 4: Distribution and disagreement in the rise probability of agents within the same team.

4.2.2 Asset Pricing Performance

In this section, we evaluate the performance of our multi-agent framework in cryptocurrency pricing. In the crypto market, the trend of an individual cryptocurrency is not limited to a binary outcome, i.e. rise or fall, but instead exists on a spectrum. Therefore, the model should also be able to explain the variation in cross-sectional cryptocurrency returns effectively. Table 2 reports the Performance comparison of out-of-sample quintile-based portfolios of our multi-agent model against baselines. We conduct two-tailed Student's t-tests for each portfolio's average return. The top 30 cryptocurrencies are sorted into quintiles based on their predicted "Rise" probability. We report the average realized weekly returns, their standard deviations, and Sharpe ratios, respectively. All portfolios are equal-weighted. "High-Minus-Low (HML)" denotes a strategy that long the "Very High" portfolio and short the "Very Low" portfolio. We observe that the HML portfolio produced by our multi-agent system not only outperforms those produced by the baseline models but also are statistically significant at the 5% level. However, all other portfolios have an insignificant average return. The performance of our multi-agent model portfolio is also better than the best-performing cryptocurrency risk factors, as illustrated in Appendix G.

The superior performance of our multi-agent model in prediction can be partially attributed to the

[2]https://platform.openai.com/docs/models

Table 2: Performance comparison of out-of-sample portfolios of our multi-agent model against baselines. *, **, and *** denote significance at the 1%, 5%, and 10% levels in two-tailed Student's t-test.

Expert agent	Portfolio	Single GPT-4o without fine-tuning			Single-agent system with fine-tuning			Ensemble single-agent system with fine-tuning			Multi-agent framework (Ours)		
		Mean	Std	Sharpe	Mean	Std	Sharpe	Mean	Std	Sharpe	Mean	Std	Sharpe
Crypto Factor	Very Low	0.0066	0.0919	0.0713	0.0053	0.0888	0.0592	0.0086	0.0772	0.1112	0.0036	0.0785	0.0464
	Low	0.0066	0.1038	0.0633	0.0039	0.0847	0.0458	0.0115	0.0925	0.1239	0.0093	0.0944	0.0983
	Medium	0.0019	0.0898	0.0210	0.0058	0.0986	0.0585	0.0090	0.0963	0.0933	0.0072	0.0850	0.0851
	High	0.0110	0.0807	0.1360	0.0180	0.0918	0.1961	-0.0001	0.0865	−0.001	0.0067	0.0934	0.0719
	Very High	0.0153	0.0843	0.1810	0.0085	0.0868	0.0978	0.0122	0.0947	0.1283	0.0144	0.0956	0.1510
	HML	0.0087	0.0630	0.1382	0.0040	0.0682	0.0592	0.0036	0.0589	0.0606	0.0108	0.0611	0.1766
		Mean	Std	Sharpe	Mean	Std	Sharpe	Mean	Std	Sharpe	Mean	Std	Sharpe
Technical	Very Low	0.0057	0.0774	0.0740	0.0053	0.0888	0.0592	0.0108	0.0787	0.1369	0.0038	0.0791	0.0475
	Low	0.0048	0.0939	0.0507	0.0039	0.0847	0.0458	0.0069	0.1026	0.0674	0.0055	0.0995	0.0553
	Medium	0.0015	0.0975	0.0157	0.0058	0.0986	0.0585	0.0049	0.0913	0.0535	0.0082	0.0850	0.0960
	High	0.0172	0.0925	0.1860	0.0180	0.0918	0.1961	0.0032	0.0829	0.0384	0.0132	0.0903	0.1461
	Very High	0.0119	0.0847	0.1407	0.0085	0.0868	0.0978	0.0152	0.0870	0.1749	0.0103	0.0869	0.1187
	HML	0.0062	0.0586	0.1057	0.0040	0.0682	0.0592	0.0044	0.0590	0.0752	0.0066	0.0524	0.1250
		Mean	Std	Sharpe	Mean	Std	Sharpe	Mean	Std	Sharpe	Mean	Std	Sharpe
Final Result	Very Low	0.0058	0.0900	0.0640	0.0053	0.0888	0.0592	0.0096	0.0776	0.1235	-0.0009	0.0792	−0.011
	Low	0.0070	0.1016	0.0688	0.0039	0.0847	0.0458	0.0086	0.1046	0.0827	0.0140	0.0958	0.1462
	Medium	0.0101	0.0974	0.1039	0.0058	0.0986	0.0585	0.0089	0.0915	0.0971	0.0041	0.0838	0.0492
	High	0.0091	0.0801	0.1131	0.0180	0.0918	0.1961	0.0046	0.0972	0.0472	0.0079	0.0951	0.0832
	Very High	0.0094	0.0849	0.1107	0.0085	0.0868	0.0978	0.0100	0.0813	0.1230	0.0160	0.0899	0.1779
	HML	0.0036	0.0695	0.0523	0.0040	0.0682	0.0592	0.0004	0.0512	0.0081	0.0169**	0.0558	0.3030

fine-tuning process. To confirm this, we visualize the distribution of the rise probabilities produced by different models, as shown in Figure 4a. We observe that the distributions of rise probabilities of individual crypto before fine-tuning exhibit a U-shaped pattern, while the distributions after fine-tuning are more centralized and align more closely with a normal or log-normal distribution. Given that the distributions of individual crypto returns are generally closer to normal or log-normal distributions, we conclude that the fine-tuning process enables the LLMs to better learn and reflect the empirical distribution of crypto returns. To evaluate the extent to which the predictions of different agents within the same group vary, we also calculate the standard deviation of the linear "Rise" probability as an indicator of disagreement. Figure 4b illustrates the level of average disagreement in both single-agent and multi-agent models. We observe that models experiencing fine-tuning exhibit lower average disagreement across the crypto team. This suggests that fine-tuning enables expert agents to better learn from historical data, avoiding random guessing. A similar experiment conducted for the market team, presented in Appendix E, yields consistent results.

4.3 Ablation Study

In this section, we evaluate the contribution of each component of our multi-agent model to portfolio performance. Table 3 presents the results of the ablation study, where key components or mechanisms are systematically removed to assess their impact on the overall portfolio performance. Annualized return of our model is calculated as $(0.8347+1)^{\frac{52}{43}} - 1 = 1.0832$. We observe that both intrateam and interteam collaboration enhance the overall portfolio performance of our multi-agent model.

Table 3: Ablation study of our multi-agent model. We evaluate the individual modules by removing each one separately and observing the changes in each metric.

Ablation	Cumulative	Mean	Std	Sharpe
Crypto Factor	0.4707▼36.4%	0.0115▼0.6%	0.0729▼0.8%	1.1395▼40.3%
Technical	0.5003▼33.4%	0.0123▼0.5%	0.0784▼0.2%	1.1354▼40.7%
News	0.7168▼11.8%	0.0160▼0.1%	0.0826▲0.2%	1.3968▼14.6%
Market Factor	0.7024▼13.2%	0.0157▼0.2%	0.0834▲0.3%	1.3576▼18.5%
Collaboration	0.8132▼2.1%	0.0166▼0.1%	0.0802▼0.0%	1.4926▼4.5%

5 Conclusion

In this work, we explored the explainable cryptocurrency investment task, a challenging problem due to the shorter history of cryptocurrencies, diverse information sources, and high market volatility compared to traditional assets. To address these challenges, we propose an explainable, multi-modal, multi-agent framework that employs multiple teams of agents collaborating both within and across teams to enable supervised learning and investment decisions across the top 30 cryptocurrencies by market capitalization. Our framework delivers an annualized return of 108.32% in portfolio performance and generates statistically significant long-short portfolio returns in asset pricing.

Limitation

A limitation of this work is that we do not evaluate the generalizability of our multi-agent model. In principle, the model can be generalized beyond the cryptocurrency domain to traditional equity markets, which offer even more abundant and diverse multimodal data.

References

Anamika Anamika and Sowmya Subramaniam. 2022. Do news headlines matter in the cryptocurrency market? *Applied Economics*.

Clifford S. Asness, Tobias J. Moskowitz, and Lasse Heje Pedersen. 2013. Value and Momentum Everywhere. *The Journal of Finance*, 68(3):929–985.

Or Biran and Kathleen McKeown. 2017. Human-centric justification of machine learning predictions. *IJCAI International Joint Conference on Artificial Intelligence*, 0:1461–1467.

Salvatore M. Carta, Sergio Consoli, Luca Piras, Alessandro Sebastian Podda, and Diego Reforgiato Recupero. 2021. Explainable Machine Learning Exploiting News and Domain-Specific Lexicon for Stock Market Forecasting. *IEEE Access*, 9:30193–30205.

CFP Board. 2022. CFP Board Issues Crypto Guidelines.

Shaen Corbet, Brian Lucey, Andrew Urquhart, and Larisa Yarovaya. 2019. Cryptocurrencies as a financial asset: A systematic analysis. *International Review of Financial Analysis*, 62:182–199.

Qianggang Ding, Haochen Shi, and Bang Liu. 2024. TradExpert: Revolutionizing Trading with Mixture of Expert LLMs.

Fan Fang, Carmine Ventre, Michail Basios, Leslie Kanthan, David Martinez-Rego, Fan Wu, and Lingbo Li. 2022. Cryptocurrency trading: a comprehensive survey. *Financial Innovation*, 8(1):1–59.

Sorouralsadat Fatemi and Yuheng Hu. 2024. FinVision: A Multi-Agent Framework for Stock Market Prediction. *Proceedings of the 5th ACM International Conference on AI in Finance*, pages 582–590.

Stéphane Goutte, Hoang Viet Le, Fei Liu, and Hans Jörg von Mettenheim. 2023. Deep learning and technical analysis in cryptocurrency market. *Finance Research Letters*, 54:103809.

Shihao Gu, Bryan Kelly, and Dacheng Xiu. 2020. Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, 33(5):2223–2273.

Andreas Hackethal, Tobin Hanspal, Dominique M. Lammer, and Kevin Rink. 2022. The Characteristics and Portfolio Behavior of Bitcoin Investors: Evidence from Indirect Cryptocurrency Investments. *Review of Finance*, 26(4):855–898.

Robert Hudson and Andrew Urquhart. 2021. Technical trading and cryptocurrencies. *Annals of Operations Research*, 297(1-2):191–220.

Liu Jing and Yuncheol Kang. 2024. Automated cryptocurrency trading approach using ensemble deep reinforcement learning: Learn to understand candlesticks. *Expert Systems with Applications*, 237:121373.

Geeta Kapur, Sridhar Manohar, Amit Mittal, Vishal Jain, and Sonal Trivedi. 2024. Cryptocurrency price fluctuation and time series analysis through candlestick pattern of bitcoin and ethereum using machine learning. *International Journal of Quality and Reliability Management*, 41(8):2055–2074.

Kelvin J.L. Koa, Yunshan Ma, Ritchie Ng, and Tat Seng Chua. 2024. Learning to Generate Explainable Stock Predictions using Self-Reflective Large Language Models. *WWW 2024 - Proceedings of the ACM Web Conference*, 12:4304–4315.

Zhizhuo Kou, Holam Yu, Jingshu Peng, Hong Kong, and China Lei Chen. 2024. Automate Strategy Finding with LLM in Quant investment.

Salim Lahmiri and Stelios Bekiros. 2019. Cryptocurrency forecasting with deep learning chaotic neural networks. *Chaos, Solitons & Fractals*, 118:35–40.

Lezhi Li, Ting-Yu Chang, and Hai Wang. 2023a. Multimodal Gen-AI for Fundamental Investment Research.

Shuqi Li, Weiheng Liao, Yuhan Chen, and Rui Yan. 2023b. PEN: Prediction-Explanation Network to Forecast Stock Price Movement with Better Explainability. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(4):5187–5194.

Yinheng Li, Shaofei Wang, Han Ding, and Hang Chen. 2023c. Large Language Models in Finance: A Survey.

Yuan Li, Bingqiao Luo, Qian Wang, Nuo Chen, Xu Liu, and Bingsheng He. 2024. A Reflective LLM-based Agent to Guide Zero-shot Cryptocurrency Trading.

Hao Liu, Carmelo Sferrazza, and Pieter Abbeel. 2023a. Chain of Hindsight Aligns Language Models with Feedback. *12th International Conference on Learning Representations, ICLR 2024*.

Xiao-Yang Liu, Guoxuan Wang, Hongyang Yang, and Daochen Zha. 2023b. FinGPT: Democratizing Internet-scale Data for Financial Large Language Models.

Yukun Liu and Aleh Tsyvinski. 2021. Risks and Returns of Cryptocurrency. *The Review of Financial Studies*, 34(6):2689–2727.

Yukun Liu, Aleh Tsyvinski, and Xi Wu. 2022. Common Risk Factors in Cryptocurrency. *The Journal of Finance*, 77(2):1133–1177.

Simona-Vasilica Oprea and Adela Bara. 2025. A Multi-agent System Based On LLM For Trading Financial Assets. *Ovidius University Annals. Economic Sciences Series*, 24(2):613–623.

Bo Pan, Jiaying Lu, Ke Wang, Li Zheng, Zhen Wen, Yingchaojie Feng, Minfeng Zhu, and Wei Chen. 2024. AgentCoord: Visually Exploring Coordination Strategy for LLM-based Multi-Agent Collaboration.

William F. Sharpe. 1994. The Sharpe Ratio. *The Journal of Portfolio Management*, 21(1):49–58.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. 2022. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *Advances in Neural Information Processing Systems*, 35.

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. 2023. AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation.

Qianqian Xie, Xiao Zhang, Weiguang Han, Yanzhao Lai, Min Peng, Alejandro Lopez-Lira, and Jimin Huang. 2023. PIXIU: A Large Language Model, Instruction Data and Evaluation Benchmark for Finance. *Advances in Neural Information Processing Systems*, 36.

Jingfeng Yang, Hongye Jin, Ruixiang Tang, Xiaotian Han, Qizhang Feng, Haoming Jiang, Shaochen Zhong, Bing Yin, and Xia Hu. 2024. Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond. *ACM Transactions on Knowledge Discovery from Data*, 18(6).

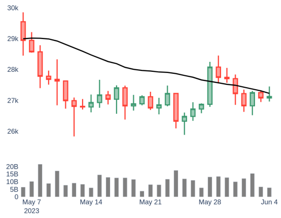
Shukang Yin, Chaoyou Fu, Sirui Zhao, Ke Li, Xing Sun, Tong Xu, and Enhong Chen. 2024. A Survey on Multimodal Large Language Models. *National Science Review*.

Yangyang Yu, Haohang Li, Yupeng Cao, Keyi Wang, Zhiyang Deng, Zhiyuan Yao, Yuechen Jiang, Dong Li, Ruey-Ling Weng, and Jordan W. Suchow. 2025. FinNLP-FNP-LLMFinLegal @ COLING 2025 Shared Task: Agent-Based Single Cryptocurrency Trading Challenge.

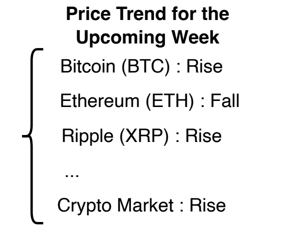
Jingyi Zhang, Jiaxing Huang, Sheng Jin, and Shijian Lu. 2024. Vision-Language Models for Vision Tasks: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(8):5625–5644.

A Data Description

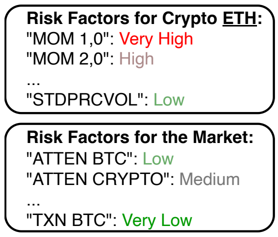
Table 4 presents the data description, associated agents, and relevant literature. The table provides an overview of our multimodal data, specifying the agent responsible for analyzing each data type and the literature fetched by the literature team to enhance the agents' explainability.



(a) 30-day candlestick charts with trading volume bars and a 30-day moving averages line.



(b) Ground truth. Binary price trends of cryptos and the whole market for the following week.



(c) Risk factors constructed from the on-chain and off-chain data.

(d) News headlines crawled from the Cointelegraph.

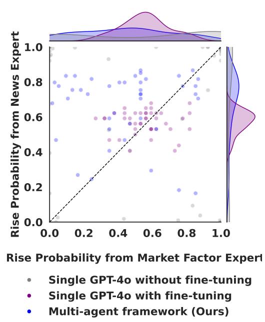
Figure 5: Multi-modal data utilized by our multi-agent framework, as described in Table 4.

B Experiments Compute Resources

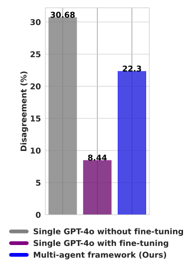
We conduct the experiment on a dual AMD Epyc 7F32 with 128 GB DDR4 ECC RAM, 2 × 240G Intel SSD, and 2 × 16 TB Seagate EXOS in Raid 1 configuration.

C Prompts and Instructions

In this section, we exhibit all system instructions and prompts of our multi-agent system.



(a) Rise probability of market team.



(b) Average disagreement of market team.

Figure 6: Distribution and disagreement in the rise probability of the market team.

Table 4: Data description and corresponding agents. We visualize the sample data in Figure 5.

Data Type	Name	Description	Agent
Chart	$v_{c_i,t}$ (Hudson and Urquhart, 2021)	30-day candlestick charts with trading volume bars and a 30-day moving averages line in week t.	Technical
Crypto Factor	$\text{MCAP}_{c_i,t}$ (Liu et al., 2022)	Log last-day market capitalization in week t.	Crypto
	$\text{PRC}_{c_i,t}$ (Liu et al., 2022)	Log last-day price in week t	
	$\text{MAXDPRC}_{c_i,t}$ (Liu et al., 2022)	Maximum price of week t.	
	MOM $1{,}0_{c_i,t}$ (Liu et al., 2022)	Past one-week return, calculated as $\frac{p_{c_i,t}}{p_{c_i,t-1}} - 1$, where p denotes the crypto price.	
	MOM $2{,}0_{c_i,t}$ (Liu et al., 2022)	Past two-week return, calculated as $\frac{p_{c_i,t-1}}{p_{c_i,t-2}} - 1$, where p denotes the crypto price.	
	MOM $3{,}0_{c_i,t}$ (Liu et al., 2022)	Past three-week return, calculated as $\frac{p_{c_i,t-2}}{p_{c_i,t-3}} - 1$, where p denotes the crypto price.	
	MOM $4{,}0_{c_i,t}$ (Liu et al., 2022)	Past four-week return, calculated as $\frac{p_{c_i,t-3}}{p_{c_i,t-4}} - 1$, where p denotes the crypto price.	
	MOM $4{,}1_{c_i,t}$ (Liu et al., 2022)	Past one-to-four-week return, calculated as $\frac{p_{c_i,t-1}}{p_{c_i,t-4}} - 1$, where p denotes the crypto price.	
	$\text{PRCVOL}_{c_i,t}$ (Liu et al., 2022)	Log average dollar volume in week t.	
	$\text{STDPRCVOL}_{c_i,t}$ (Liu et al., 2022)	Log standard deviation of dollar volume in week t.	
Market Factor	ATTN BTC_t (Liu and Tsyvinski, 2021)	Google search data for the word Bitcoin minus its average of the previous four weeks (calculated as $\Delta a_t^{\text{BTC}} = a_t^{\text{BTC}} - \frac{1}{4}\sum_{i=1}^{4} a_{t-i}^{\text{BTC}}$), and then normalized all historical Δa^{BTC} to have a mean of zero and a standard deviation of one to get the normalized Δa_t^{BTC}.	Market
	ATTN CRYPTO_t (Liu and Tsyvinski, 2021)	Google search data for the word cryptocurrency minus its average of the previous four weeks(calculated as $\Delta a_t^{\text{CRYPTO}} = a_t^{\text{CRYPTO}} - \frac{1}{4}\sum_{i=1}^{4} a_{t-i}^{\text{CRYPTO}}$), and then normalized all historical Δa^{CRYPTO} to have a mean of zero and a standard deviation of one to get the normalized $\Delta a_t^{\text{CRYPTO}}$.	
	UNI ADDR_t (Liu and Tsyvinski, 2021)	The first difference between the Bitcoin wallet in week t and $t-1$.	
	ACT ADDR_t (Liu and Tsyvinski, 2021)	The first difference between the Active Bitcoin addresses in week t and $t-1$..	
	TXN_t (Liu and Tsyvinski, 2021)	The first difference between Bitcoin transactions in week t and $t-1$..	
	PAY_t (Liu and Tsyvinski, 2021)	The first difference the Bitcoin payments in week t and $t-1$..	
Text	N_t (Anamika and Subramaniam, 2022)	News headlines from Cointelegraph in week t.	News

System Instruction 1: Explanation Team.

You are a professional cryptocurrency analyst, specializing in explaining the predicted target based on the provided knowledge and information. You should internalize the provided knowledge to generate a comprehensive explanation without explicitly referring to the literature. Your output should be in a single paragraph.

Prompt 1: Market Factor Analyst.

Learn the following cryptocurrency investment knowledge. Using this knowledge, explain the predicted target for the upcoming week based on the provided information. The market factors have been categorized into Very High, High, Medium, Low, and Very Low using the first two years of data. The predicted market return has been categorized into Rise or Fall.
Knowledge: *{literature}* (End of knowledge)
Information: *{factors/news}* (End of information)
Market trend: *{future market trend}* (End of market trend)

D Explanation Performance Metrics

In the context of crypto portfolio management, we define model explainability as the ability to generate rationales for cryptocurrency and market trend predictions grounded in professional asset pricing knowledge from the field of finance. Figure 7 compares the example outputs of our crypto factor expert agent with fine-tuning and GPT-4o without fine-tuning. We observe that the explanation generated by the expert agent after fine-tuning incorporates significantly more asset pricing terminologies from the provided literature. In addition, we use the following metrics generated by the GPT-4o to evaluate the model's explainability:

- **Professionalism:** Does the explanation reflect expertise and professionalism in the field of finance?
- **Objectivity:** Is the explanation presented in an unbiased and neutral manner?

Prompt 2: News Analyst.

Learn the following cryptocurrency investment knowledge. Using this knowledge, explain the predicted target for the upcoming week on the provided news headlines. The predicted market return has been categorized into Rise or Fall.
Knowledge: {*literature*} (End of knowledge)
Information: {*factors/news*} (End of information)
Market trend: {*future market trend*} (End of market trend)

Prompt 3: Crypto Factor Analyst.

Learn the following cryptocurrency investment knowledge. Using this knowledge, explain the predicted price trend of {*target crypto*} for the upcoming week based on the provided information. The data for the top 30 cryptocurrencies, including {*target crypto*}, have been categorized into Very High, High, Medium, Low, and Very Low. Their respective predicted price trend has been categorized into Rise or Fall.
Knowledge: {*literature*} (End of knowledge)
Information: {*factors*} (End of information)
Price trend: {*future price trend*} (End of price trend)

- **Clarity & Coherence:** Is the explanation easy to understand, and does it follow a logical structure that connects different factors effectively?
- **Consistency:** Does the explanation align with the provided data and avoid contradictions?
- **Rationale:** Does the explanation provide a detailed reasoning process that clearly articulates how the metrics influence performance?

Figure 8 reports the average score for each metric. From Figure 8, we observe that our multi-agent model outperforms the single-agent model with fine-tuning across all metrics. This highlights the advantage of the multi-agent framework, where each domain-specific expert agent, after training, can generate more accurate and specialized explanations compared to a single generalized agent.

Prompt 4: Technical Analyst.

Text:
Learn the following cryptocurrency investment knowledge. Using this knowledge, explain the predicted price trend of {*target crypto*} for the upcoming week based on the provided candlestick chart. The chart includes candlesticks that depict daily opening, high, low, and closing prices. It then overlays a 30-day moving average closing price. The bottom of the chart shows daily trading volume.
Knowledge: {*literature*} (End of knowledge)
Price trend: {*future price trend*} (End of price trend)
Image URL: {*URL*}

System Instruction 2: Fine-Tuning Market and Crypto Team.

You are a professional cryptocurrency analyst, specializing in predicting next week's {"price trend of a cryptocurrency"/"market trend"} based on the provided information.
Your output should be in the form of:
Target: (predicted target)
Explanation: (your explanation)

E Effect of Fine-Tuning on Prediction Distribution and Agent Disagreement

We also visualize the distribution of the rise probabilities of crypto market extracted from the outputs of the single GPT-4o without fine-tuning, single GPT-4o with fine-tuning, and our multi-agent model, as shown in Figure 6a. We observe that the distributions of rise probabilities of market before fine-tuning exhibit a flat pattern, while the distributions after fine-tuning are more centralized and align more closely with a normal or log-normal distribution. Given that the distributions of individual market returns are generally closer to normal or log-normal distributions, we conclude that the fine-tuning process enables the LLMs to better learn and reflect the empirical distribution of crypto returns. To evaluate the extent to which the predictions of different agents within the same group vary, we also calculate the standard deviation of the linear "Rise" probability as an indicator of disagreement. Fig-

Prompt 5: Fine-Tuning Market and Crypto Team.

User:
Analyze the following information of crypto to determine its target in a week. Please respond with Rise or Fall and provide your reasoning for the prediction.:
{*info*} (End of information)
Assistant:
{“Price trend"/“Market trend"}: {*trend*}
Explanation: {*expanation*}

System Instruction 3: Market and Crypto Team Prediction.

You are a professional cryptocurrency analyst, specializing in predicting next week’s {“price trend of a cryptocurrency"/“market trend"} based on the provided candlestick chart. Your output should be in the form of:
{“Price trend"/“Market trend"}: (predicted target)
Explanation: (your explanation)

ure 6b illustrates the level of average disagreement in our multi-agent model compared to the baseline models. We observe that models experiencing fine-tuning exhibit lower average disagreement across the market team. This suggests that fine-tuning enables expert agents to better learn from historical data, avoiding random guessing.

F Portfolio Performance Evaluation Metrics

In this section, we define all metrics used to evaluate the portfolios:

Prompt 6: Market and Crypto Team Prediction.

Analyze the following {“cryptocurrency"/"market"} information to determine the strength of the {“price trend of a cryptocurrency"/“market trend"} in a week. Please respond with Rise or Fall and provide your reasoning for the prediction.
Information: {*factors/news/charts*} (End of information)

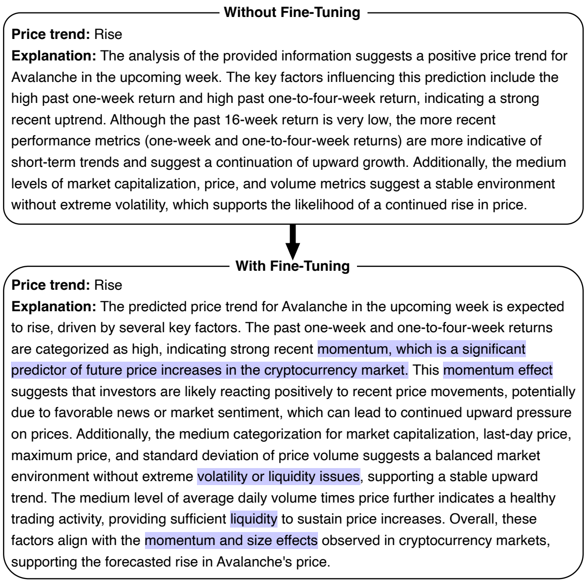


Figure 7: A comparison of the example outputs of our crypto factor expert agent with fine-tuning and GPT-4o without fine-tuning. We highlight the asset pricing terminologies that the model has learned from (Liu et al., 2022).

- **Cumulative Return (Cumulative)** (Asness et al., 2013) measures the total changes in the price of a portfolio over the trading period, calculated as $\prod_{t=1}^{T}(1+r_t)-1$, where T denotes the total number of weeks over the trading period and r_t denotes the weekly return.

- **Weekly Return Mean (Mean)** (Gu et al., 2020) measures the average of weekly returns over the trading period, indicating the portfolio’s typical weekly performance.

- **Weekly Return Standard Deviation (Std)** (Gu et al., 2020) measures the standard deviation of

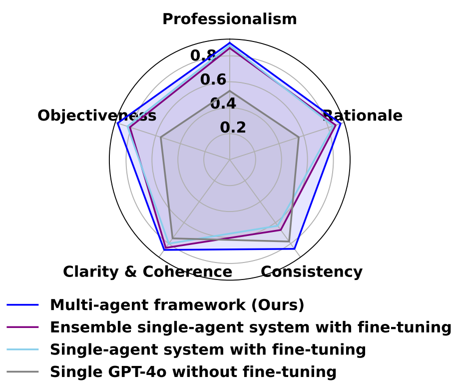


Figure 8: Comparison of explanation quality from our multi-agent model against baselines.

weekly return over the trading period, representing the volatility of the portfolio.

- **Sharpe Ratio (Sharpe)** (Sharpe, 1994) measures the risk-adjusted return, calculated as $\frac{r_t - r_f}{\sigma_t}$, where r_t denotes the weekly return mean, r_f is the risk-free rate, σ denotes the weekly return standard deviation.

G Asset Pricing Performance of Cryptocurrency Risk Factors

Table 5 reports the performance comparison of out-of-sample quintile-based portfolios of the best-performing cryptocurrency risk factors.

Table 5: Performance of out-of-sample portfolios of best-performing cryptocurrency risk factors.

Factor	Portfolio	MOM 1,0			MOM 4,0			MOM 4,1		
		Mean	Std	Sharpe	Mean	Std	Sharpe	Mean	Std	Sharpe
Top Factor	Very Low	-0.0043	0.0853	−0.0499	-0.0014	0.0888	−0.0157	0.0008	0.0838	0.0101
	Low	0.0082	0.0850	0.0961	0.0111	0.1012	0.1098	0.0136	0.1055	0.1291
	Medium	0.0152	0.1071	0.1421	0.0148	0.0952	0.1558	0.0119	0.0959	0.1242
	High	0.0117	0.0862	0.1359	0.0117	0.0826	0.1414	0.0123	0.0915	0.1340
	Very High	0.0103	0.0872	0.1180	0.0051	0.0877	0.0578	0.0029	0.0810	0.0359
	HML	0.0146	0.0677	0.2156	0.0073	0.0618	0.1180	0.0027	0.0516	0.0524