$$\pi_{\theta^*} = \arg\max_{\pi_\theta} \mathbb{E}_{\pi_\theta}[\sum_{t=0}^{T} \gamma(r_t) \mid s_t = s, \mu_t = \mu], \tag{1}$$

where r_t is a reward from real-world trading, γ is the discount factor and the state s_t includes market indicators and other relevant financial data. Actions a_t are determined by:

$$\pi(a_t \mid s_t, \mu_t) \equiv \mathcal{D}(LLM(\phi_D(s_t, \mathcal{M}_{ret}, \mu_t))), \tag{2}$$

where μ_t is the trading strategy. The prompt template $\phi_D(\cdot)$ is meticulously designed, and fed into an LLM, then the response from the LLM is processed by the parsing function $\mathcal{D}(\cdot)$ to obtain an action. The policy π_θ is updated using gradient ascent. Finally, we update the $\mathcal{M}_I$ in real time. The all market information (e.g., stock prices), the summarized query Q_t, and reflections on the action are all continuously updated in $\mathcal{M}_I$ to guide future decisions. A simplified prompt is as follows[4],

<Prompt Template>
You are {Dave Profile}. The market environment today includes {Prices}, {News}. Through financial analysis tools, {Tool Results} can be obtained. The output format should be JSON, such as {Examples}.

3.4 Coordination of Multi-Agents

To simulate real-world fund companies, functions such as simulated trading and market reports are integrated into our framework with collaboration among the four agents. In this section, we will provide a detailed introduction to the three types of collaborative meetings, followed by the decision-making process employed by manager Otto.

3.4.1 Market Analysis Meeting

The Market Analysis Meeting, scheduled weekly following the last trading day, integrates the expertise of Emily, Bob, and Dave, to generate a comprehensive market report covering news analysis, industry analysis, individual stock analysis and more. The collaborative approach ensures a balanced perspective, combining qualitative insights with quantitative rigor.

- Emily first uses the FinReport tool (Li et al., 2024) to conduct an overall analysis based on historical data and news inputs, encompassing economic indicators, global trends, and geopolitical influences.

[4]The templates will change based on different tasks indicated by $\phi_{(\cdot)}$, fully disclosed in the Appendix.

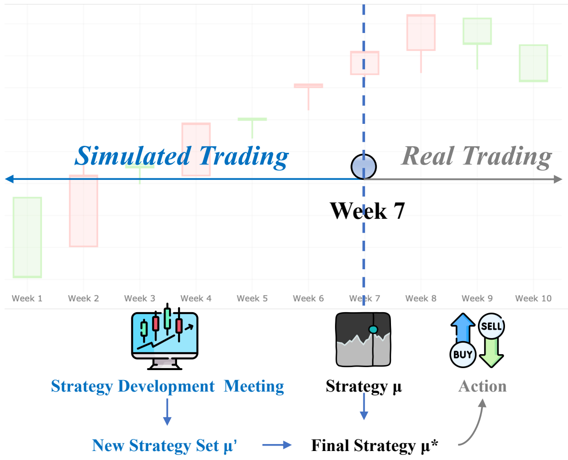

Figure 3: Our framework has been optimized to obtain rewards from both simulated and real-world trading.

- Bob will then provide a quantitative analysis based on Emily's report, utilizing statistical tools including trend forecasting (Chaudhari and Thakkar, 2023), factor analysis (Fama and French, 2015), and other relevant methods. The analysis will incorporate historical data and news inputs to evaluate market trends, industry performance, and individual stock metrics.

- Dave will deliver a risk analysis using the Volatility Assessment Tool (Mieg, 2022), focusing on market volatility, risks within specific industries, and vulnerabilities of individual stocks.

The conclusions drawn by the three analysts will be fed into the LLM in a prompt template $\phi_R(\cdot)$, resulting in a comprehensive market report, which is then stored in the Reports Memory $\mathcal{M}_R$ for future decision.

3.4.2 Strategy Development Meeting

The strategy analysis meeting is held weekly after the last trading day. This meeting is crucial for implementing simulated trading to test new strategies $\mu_t^{'}$. During this meeting, Bob is responsible for testing new strategies, while Emily and Dave provide advice on market conditions and risk management. This meeting ensures that all new strategies are rigorously evaluated and refined before deployment.

- Bob will conduct simulated trading to test all potential new strategies, and outline their characteristics along with possible optimization approaches. As depicted in Figure 3, Bob will utilize statistical tools such as the Simulation Optimization Toolkit (Ha and Mueller, 2024) and the Strategy Analysis Suite (Gaikwad et al., 2012), along with all the data prior to the current time (e.g. Week 7) to undertake a simulated trading backtest. Bob will select the strategies that perform best in the historical data to form a new strategy set μ' [5] .

- Dave performs a risk analysis on the new strategy using the RiskAnalyzer toolkit (Yang et al., 2020) , and suggests optimization directions from risk management.
- Emily conducts market analysis on the new strategy and proposes optimization directions based on market events.

Finally, the three analysts consolidate their findings, encompassing the detailed characteristics and risk assessments of the new strategy μ_{t+1}, into the strategy memory $\mathcal{M}_S$ to aid action decisions.

3.4.3 Risk Alert Meeting

The risk alert meeting is triggered by a risk threshold to mitigate investment risk, which is defined as the combination of Portfolio Beta (β_p), Liquidity Ratio (LR), Sector Exposure (SE_j), and Volatility (σ_p) as follows,

$$R_{score} = w_1\beta_p + w_2(\frac{1}{LR}) + w_3 \max(SE_j) + w_4\sigma_p \quad (3)$$

where w_i are risk factor weights. This meeting will be triggered once $R_{score} > 0.75$.

- The Risk Score Assessment tool (Alexander, 2008) is used by Dave to get a comprehensive risk analysis, including portfolio Beta value, Value at Risk (VaR), and sector concentration.
- Bob conducts quantitative analysis using the StressTestPro tool (Koliai, 2016), which takes historical data and stress scenarios as inputs and outputs a risk severity score $\eta \in [0, 1]$, helping to evaluate potential market impacts.
- Emily uses the SentimentAnalyzer tool (Araci, 2019) to analyze market sentiment $\tau \in [-1, 1]$ for high-risk assets. This tool processes financial news and social media data to generate a sentiment score reflecting the market's positive, neutral, or negative stance.

[5] We construct a strategy pool consisting of multiple index permutations, detailed in the Supplementary Material.

Ultimately, the conclusions drawn by the three agents mentioned above are integrated by Otto to make decisions regarding high-risk situations. The modified policy $\pi_{\theta^*}^{risk}$ is:

$$\pi_{\theta^*}^{risk} = \arg\max_{\pi_\theta} \mathbb{E}_{\pi_\theta}[\sum_{t=0}^{T} \gamma((1-\lambda)r_t + \lambda r_t^{risk})], \quad (4)$$

where λ is a risk adjustment factor, and the risk-based reward is $r_t^{risk} = f(R_{score}, \eta, \tau)$. This approach ensures that our QuantAgents remains responsive to urgent risk situations while maintaining its long-term learning capabilities.

3.4.4 Decision Making from Otto.

Manager Otto synthesizes all information and executes trading actions according to the optimal strategy. With the introduction of simulated trading, Otto now receives rewards from two distinct sources: real- world trading and the simulated trading of strategies. As depicted in Figure 3, Otto conducts the action in accordance with the policy μ at the present moment (e.g., Week 7), along with the newly available policy μ', as the final policy μ^*.

Consequently, Eqn. 1 will be revised as follows:

$$\pi_{\theta^*} = \arg\max_{\pi_\theta} \mathbb{E}_{\pi_\theta}\left[\sum_{t=0}^{T} \gamma\big(w_t^{sim} r_t^{sim} + w_t^{real} r_t^{real}\big)\right] \quad (5)$$

where r_t^{sim} is the reward from the simulated trading, and r_t^{real} is the reward from the real-world trading environment. w_t^{sim} and w_t^{real} are adaptive weights. The adaptive weights are updated based on the relative performance:

$$w_t^{sim} = \sigma\left(\frac{\sum_{i=t-n}^{t} r_i^{sim}}{\sum_{i=t-n}^{t}(r_i^{sim} + r_i^{real})}\right), \quad w_t^{real} = 1 - w_t^{sim} \quad (6)$$

where $\sigma(\cdot)$ is the sigmoid function and n is the number of recent time steps considered.

4 Experiments

4.1 Datasets

The constituents of the NASDAQ-100 from January 1, 2010, to December 31, 2023, will serve as the

Table 2: Cumulative Returns Comparison of our QuantAgents and all baselines. **Bold** represents optimal performance, while underline represents suboptimal.

Categories	Models	ARR(%)	TR(%)	SR	CR	SoR	MDD(%)	VoL(%)	ENT	ENB
Market Index	NDX	9.84	32.52	0.64	1.38	13.07	35.58	1.52	—	—
Classical	MV	11.3	37.87	0.72	3.27	22.05	64.15	5.79	1.01	1.02
	ZMR	4.19	13.1	0.63	2.52	18.43	72.89	5.82	1.43	1.09
	TSM	5.68	18.02	0.64	3.11	17.27	58.36	5.65	1.03	1.07
RL-based	SAC	22.14	82.23	0.84	2.99	23.63	40.13	2.85	1.49	1.11
	DeepTrader	32.06	130.29	1.27	7.16	30.31	29.16	2.81	1.88	1.19
	AlphaMix+	32.51	132.72	1.49	5.76	30.66	40.71	2.85	2.76	1.36
LLM-based	FinGPT	36.71	155.52	1.66	6.34	42.31	37.99	2.83	1.94	1.21
	FinMem	37.73	161.25	1.89	6.16	43.02	40.19	2.82	2.25	1.24
	FinAgent	45.31	206.83	2.25	6.98	47.66	38.48	2.92	2.71	1.38
	HedgeAgents	49.25	230.39	2.41	6.53	45.21	23.65	1.99	2.68	1.35
Ours	**QuantAgents**	**58.68**	**299.55**	**3.11**	**11.38**	**66.94**	**16.86**	**1.43**	**2.97**	**1.49**
Improvement(%)		19.15	30.02	29.05	58.94	40.45	28.71	5.92	7.61	7.97

evaluation dataset. This dataset includes daily trading data for each stock, comprising open, high, low, and closing prices, as well as trading volume, along with 60 standard technical indicators for analysis. Moreover, we incorporated daily news updates, company financial reports, and macroeconomic policy information for each asset. These data were sourced from Yahoo Finance for market data and the Alpaca News API for texts.

4.2 Evaluation Metrics

We compare all models in terms of 9 financial metrics following (Sun et al., 2023b; Qin et al., 2023), which include 2 profit metrics: Total Return (TR), Annual Return Rate (ARR); 3 risk-adjusted profit metrics: Sharpe Ratio (SR), Calmar Ratio (CR), Sortino Ratio (SOR); 2 risk metrics: Maximum Drawdown (MDD), Volatility (VOl); and 2 diversity metrics: Entropy (ENT) and Effect Number of Bets (ENB). Higher values are preferred for all metrics except MDD and VOL, where lower values indicate better performance.

4.3 Implementation Details

The dataset will be temporal split, with data from January 1, 2010, to December 31, 2020, constituting the training set, and data from January 1, 2021, to December 31, 2023, serving as the test set. For LLM-based approaches like QuantAgents, we adopt "GPT-4o-2024-05-13" as the foundation model, setting the temperature to 0.7 to balance consistency and creativity. The memory module operates as a similarity-based storage and retrieval system, utilizing the text-embedding-3-large model(OpenAI, 2023) for text vectorization. The retrieval process is configured to return the top 10 results for efficient information recall.

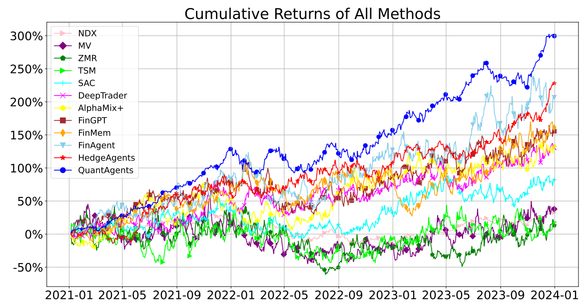

Figure 4: Cumulative Returns Comparison.

4.4 Overall Performance Comparison

Table 2 presents a comprehensive comparison of QuantAgents against a diverse set of baseline models. These baselines include 1) three classical rule-based quantitative investment strategies: MV(Yu and Yuan, 2011), ZMR(Eeckhoudt and Laeven, 2018), and TSM(Moskowitz et al., 2012); 2) three reinforcement learning-based financial agents: SAC (Haarnoja et al., 2018), DeepTrader(Wang et al., 2021b), and AlphaMix+(Sun et al., 2023a); 3) and four LLM-based methods: FinGPT(Yang et al., 2023a), FinMem(Yu et al., 2023), FinAgent(Zhang et al., 2024) and HedgeAgents(Li et al., 2025). The following observations can be made:

1) RL-based methods outperform rule-based strategies in both profitability and risk-adjusted performance. For instance, DeepTrader achieves an annualized return rate (ARR) of 32.06% and a Sharpe ratio (SR) of 1.27, surpassing the best-performing classical strategy (MV) with an ARR of 11.3% and SR of 0.72.

2) LLM-based methods further improve upon RL-based approaches, showcasing the power of