causal structures from complex semantic information rather than performing high-frequency precise numerical optimization. This capability gives them greater potential in mid-to-long-term strategy formulation. Under the *Single-Step Trading Task* setup, the agent primarily focuses on the price direction of the next step, which somewhat forces the LLM struggling to build coherent strategies or cognition. By introducing position continuity and strategy coherence, the *Position-Aware Trading Task* provides LLM agents with a more fitting testing platform, fully unlocking their potential in long-term decision-making and complex information integration. Furthermore, this setup enables agents to construct an internally consistent position management logic centered around fundamentals, market sentiment, and macro semantic information. The combination of long-term and semantic-driven strategies is precisely where LLMs hold their advantages in financial trading tasks.

4 Architecture of FinPos

The architecture of FinPos (Fig. 2) consists of three core modules. The **market signal processing module** adopts a two-level agent hierarchy, consisting of filtering agents and analyst agents, to distill raw market data into structured, decision-relevant signals. The **trading decision module** employs a dual decision framework that explicitly decouples quantity decision-making from directional determination. The **multi-timescale reward module** adopts a multi-timescale reward feedback mechanism to experientially internalize the long-term risk and return implications of position changes. The details of these three modules are presented in the subsequent sections, while all agent prompts are provided in the Appendix A.

4.1 Market Signal Processing and Analysis

FinPos draws inspiration from institutional investment workflows. It assigns specialized agents to distinct information domains forming a division of labor similar to that in private equity firms. For each domain, we further establish a two-tier structure: Signal Processing Agents and Analysis Agents.

Signal Processing Agents: These agents are responsible for preprocessing and filtering high-noise, heterogeneous market data streams. Through domain-specific prompts and heuristic rules, the agents clean, compress, and prioritize information by relevance and importance. For example, large volumes of low-value or weakly related news are downweighted or discarded, while impactful macroeconomic policies or firm-level events are highlighted and passed on to subsequent modules.

Analysis Agents: These agents analysis the filtered data. We observe that large language models often exhibit financial hallucinations in market scenarios, making it difficult for them to accurately capture the key factors influencing stock prices. They tend to rely solely on explicit keywords in news while overlooking the underlying market logic. For example, an LLM may fail to recognize that news about the S&P 500 is strongly correlated with the performance of leading U.S. stocks. To mitigate this issue, we explicitly inject financial knowledge into the prompts, thereby strengthening their causal awareness. As financial reasoning is gradually integrated into the prompts, the agents evolve from mechanical summarizers into analysts with genuine financial reasoning ability, capable of generating deeper insights and producing more accurate interpretations of market dynamics.

Subsequently, as illustrated in Fig. 2, all analytical results are aggregated into a Hierarchical Memory Module. In this module, important long-term information (e.g., annual reports) is allocated to deep memory, while volatile short-term information (e.g., corporate news) is stored in shallow memory. The hierarchy of memory is not static but dynamically adjusted through post-decision reflection: memories that repeatedly prove their validity are gradually migrated into deeper layers, thereby increasing their weight in future decision-making.

4.2 Dual Trading Decision

In real-world trading, an account's current holdings directly reflect its risk exposure. Professional traders adjust their buy–sell decisions dynamically based on existing positions to maintain a relative balance between risk and return. FinPos adopts a dual-agent framework that embeds position awareness into both **the Direction Decision Agent** and **the Quantity and Risk Decision Agent**. By integrating multi-source information retrieved from the memory layer, the dual-agent framework enables a decision-making mechanism that more closely mirrors practical trading logic.

4.2.1 The Direction Decision Agent

Incorporating position awareness requires integrating longer-term information. The direction decision agent combines structured memory (news, fi-

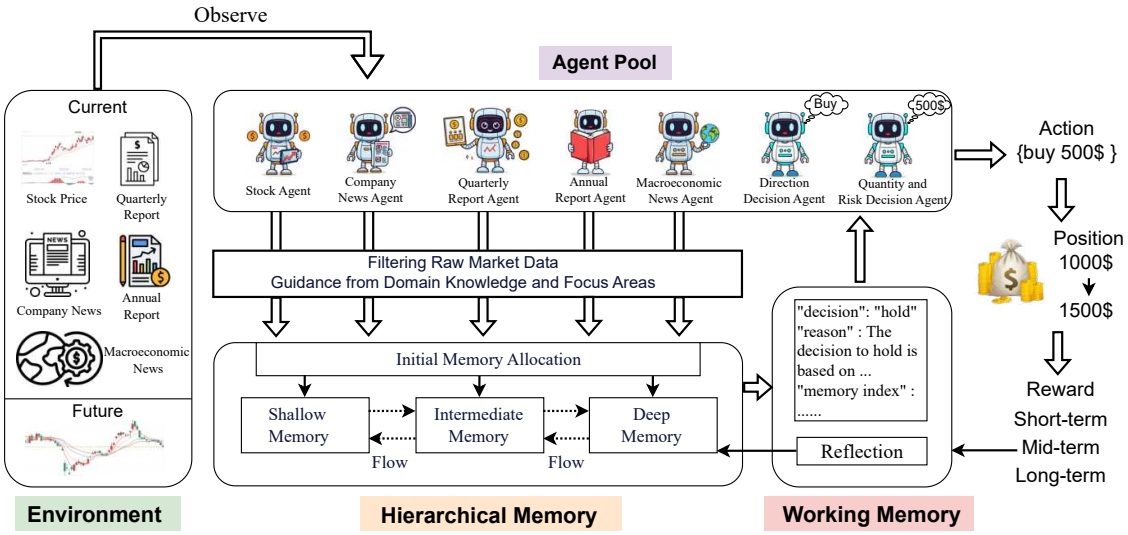

Figure 2: **Architectural Details of FinPos:** Initially, multiple analysis agents leverage domain knowledge to gather diverse information from the environment, subsequently storing it in the memory module. The memory module utilizes a memory allocator to distribute the acquired information across memory layers of varying depths. Subsequently, the most pertinent information for the current decision is placed into working memory, where dual decision agents generate trading actions, while multi-timescale reward signals support reflective updates that consolidate experiential knowledge into deeper memory layers.

nancial reports, and past actions) with the current portfolio state and explicitly specifies the strategic intent of each action—whether it represents a long-term position-building move or a short-term tactical adjustment to exploit local trends. This explicit articulation not only enhances interpretability for human supervisors but also provides downstream quantity agents with a clear strategic context. To prevent the agent from over-relying on short-term fluctuations, we design prompts that encourage it to factor in longer horizons, thereby gradually cultivating its long-term planning capability.

4.2.2 Quantity and Risk Decision Agent

FinPos introduces the Quantity and Risk Decision Agent, which determines the specific trade size after a directional decision has been made. This agent determines trade sizes through an explicit risk-aware mechanism that jointly incorporates the current position state, structured memory, and CVaR-based risk references (Conditional Value at Risk (Rockafellar et al., 2000); see Appendix B.2), with per-step transaction sizes capped by the 95% CVaR to control exposure under volatility. This design ensures that every trade size is grounded in the agent's current position and risk tolerance, making exposure control an explicit, first-class component of FinPos's decision process.

4.3 Multi-Timescale Reward Design

To guide the agent toward non-myopic trading behavior, we design a trend-aware reward function that leverages *multi-timescale market signals*. Specifically, we calculate three future price trends in each timestep t. $M_t^s = price[t+1] - price[t]$: 1-day trend (short-term), $M_t^m = price[t+7] - price[t]$: 7-day trend (mid-term), $M_t^l = price[t+30] - price[t]$: 30-day trend (long-term). We define the multi-timescale score M_t as:

$$M_t = M_t^s + M_t^m + M_t^l \tag{3}$$

This score represents the aggregated expected trend across multiple timescale (e.g., 1-day, 7-day, and 30-day). The reward at time t, $Reward_t$ is defined as follows:

$$Reward_t = \begin{cases} -(M_t)^2, & pos_t = pos_{t-1} \\ pos_t \times M_t, & otherwise \end{cases} \tag{4}$$

$$pos_t = pos_{t-1} + d_t \times q_t \tag{5}$$

d_t and q_t denote the direction and quantity of the purchasing decision at the current time step, respectively. Crucially, this reward is used only during training-time simulation, no future-dependent signals are accessed at test time.

It serves two complementary roles in guiding the dual-agent architecture: **(1) Guiding the Direction Agent:** The term $\mathsf{pos}_t \times M_t$ provides positive

reinforcement when the agent's directional decision (`buy`/`sell`/`hold`) aligns with the multi-horizon trend M_t. This encourages the Direction Agent to move beyond short-term noise and base decisions on longer-term market signals. **(2) Guiding the Quantity Agent:** During LLM reflection, the Quantity Agent uses this reward to learn when to adjust exposure. Through this process, the agent internalizes the concept of position as a dynamic state. In our experiments, we find that during periods of high volatility, agents tend to maintain a inactive state. To address this, we introduce a quadratic penalty on $|M_t|$. whenever the position remains unchanged. This discourages missed opportunities in volatile markets while also preventing excessive trading in stable periods.

5 Experiments

5.1 Experimental Setup

5.1.1 Datasets

Our research utilizes actual financial data sourced from publicly authoritative providers and encompasses various forms of market information. (1) Stock prices, obtained from Yahoo Finance, include daily open-high-low-close-volume (OHLCV) data. (2) Company news, retrieved using the Finnhub stock API, includes articles' related company name, publication date, headline, and summary text, processed for sentiment and semantic relevance. (3) Macroeconomic news, also retrieved via the Finnhub stock API, contains publication dates, headlines, and summary texts for each article. (4) 10-Q (quarterly reports) and 10-K (annual reports), accessed through the U.S. Securities and Exchange Commission (SEC) EDGAR API. These documents provide financial information that publicly listed companies are required to disclose, and standardized into a daily time series format.

5.1.2 Comparative Methods

To assess the effectiveness of FinPos, we conduct comparative experiments with four LLM-based trading agents: FinGPT (Yang et al., 2023), FINMEM (Yu et al., 2024a), FinAgent (Zhang et al., 2024), and FINCON (Yu et al., 2024b); three deep reinforcement learning (DRL) methods (A2C, PPO, DQN); two rule-based methods (MACD, RSI); and two market baselines (Random). All compared models are evaluated using identical data splits and executed under the same market conditions. Additional details are provided in Appendix C.

5.1.3 Evaluation Metrics

We evaluate performance using cumulative return (CR%), sharpe ratio (SR), and maximum drawdown (MDD%). CR reflects overall profitability, while SR measures the risk-adjusted return by showing how much excess return it generates per unit of total risk. MDD quantifies the worst-case loss over the trading horizon, which in the position-aware trading task directly reflects exposure and vulnerability under held positions. CR has been introduced earlier (see Eq. (2)), and the formulas for SR and MDD are provided in Appendix B.1.

5.1.4 Implementation Details

All LLM trading agents are deployed using GPT-4o (Hurst et al., 2024) with a temperature setting of 0.7. All models are evaluated under the position-aware trading task formally defined in Sec. 3.2, ensuring profits and risks are calculated with explicit holding dynamics. Training covered Jan 2024–Feb 2025; testing spanned Mar–Sep 2025, covering the U.S. election period and other major macroeconomic events, making it a representative and challenging evaluation setting.

5.2 Main Results

We evaluated FinPos on five representative stocks: TSLA, AAPL, AMZN, NFLX, and COIN. As shown in Tab. 1, FinPos consistently outperforms all baseline methods, achieving higher CRs, SRs, and lower MDDs across all assets. On high-volatility stocks such as TSLA, AAPL, and COIN, FinPos exhibits strong robustness under turbulent market conditions. While DRL baselines (A3C, DQN, PPO) and LLM-based agents (FinGPT, FinAgent) suffer substantial losses and severe drawdowns (e.g., A3C loses over 80% on TSLA with an MDD of 96.6%), FinPos consistently maintains positive returns, achieving CRs of 62.15% and 36.31% on TSLA and AAPL, respectively, with well-controlled downside risk. For assets with clearer trends, such as AMZN and NFLX, FinPos continues to deliver stable and risk-adjusted performance. FinPos achieves a CR of 30.35% on AMZN with an MDD of 18.44%, and a CR of 28.65% on NFLX with a high SR of 1.02, outperforming all baseline agents in both return and risk control.

5.3 Ablation Studies

We select TSLA, AAPL, and AMZN as test assets due to their rich coverage of news articles, earnings events, and macroeconomic sensitivities. We