
P1GPT: A MULTI-AGENT LLM WORKFLOW MODULE FOR MULTI-MODAL FINANCIAL INFORMATION ANALYSIS

Chen-Che Lu

Neurowatt AI

Intelligent Technology Research and Development

peter@neurowatt.ai

Yun-Cheng Chou

Neurowatt AI

Intelligent Technology Research and Development

oro@neurowatt.ai

Teng-Ruei Chen

Neurowatt AI

Intelligent Technology Research and Development

luka@neurowatt.ai

ABSTRACT

Recent advances in large language models (LLMs) have enabled multi-agent reasoning systems capable of collaborative decision-making. However, in financial analysis, most frameworks remain narrowly focused on either isolated single-agent predictors or loosely connected analyst ensembles, and they lack a coherent reasoning workflow that unifies diverse data modalities. We introduce **P1GPT**, a layered multi-agent LLM framework for multi-modal financial information analysis and interpretable trading decision support. Unlike prior systems that emulate trading teams through role simulation, P1GPT implements a structured reasoning pipeline that systematically fuses technical, fundamental, and news-based insights through coordinated agent communication and integration-time synthesis. Backtesting on multi-modal datasets across major U.S. equities demonstrates that P1GPT achieves superior cumulative and risk-adjusted returns, maintains low drawdowns, and provides transparent causal rationales. These findings suggest that structured reasoning workflows, rather than agent role imitation, offer a scalable path toward explainable and trustworthy financial AI systems.

Keywords Multi-agent system · Large language model · Financial trading · Workflow · Multi-modal data · Explainability

1 Introduction

1.1 Background and Motivation

Financial markets are inherently complex systems, where asset prices are driven by a confluence of heterogeneous signals such as corporate fundamentals, market microstructure patterns, global macroeconomic indicators, and unstructured textual data from news and social media. Traditional quantitative models and rule-based trading systems have achieved partial success, but they often rely on restrictive assumptions and struggle to adapt to rapidly evolving market conditions. In parallel, large language models (LLMs) have recently demonstrated remarkable progress in natural language understanding and domain adaptation. Notably, domain-specific initiatives such as BloombergGPT [1] and the open-source FinGPT [2] illustrate how LLMs trained on financial corpora can deliver state-of-the-art performance in tasks like sentiment classification and question answering, lowering the barrier for financial AI research. These developments suggest that LLMs could play a pivotal role as intelligent assistants in financial analysis and trading.

1.2 Challenges in Financial Trading

Despite promising progress, deploying LLMs in real-world trading scenarios remains challenging. First, most existing approaches are single-agent systems focused on narrow tasks such as stock prediction or sentiment extraction [3],

which limits their ability to capture the collaborative, multi-role workflows observed in professional investment firms [4]. Second, inter-agent communication is often unstructured, typically relying on free-form natural language exchanges that can introduce ambiguity and hinder consistent reasoning [5]. Third, current LLM-driven trading systems emphasize predictive accuracy but provide limited explainability, making them difficult to integrate into regulated financial environments where transparency and risk control are essential [6]. Finally, scalability is a persistent issue, as highlighted by large-scale simulation studies such as MASS [7], which underscore the difficulty of coordinating many agents and fusing multi-modal signals in complex markets.

1.3 Contributions of P1GPT

This work introduces **P1GPT**, a modular multi-agent LLM workflow designed to address the above challenges in financial information analysis and trading. The framework is inspired by the organizational structure of real-world trading firms and is organized into four layers: an *Input Layer* for parsing heterogeneous financial and textual data, a *Planning Layer* for task decomposition and agent assignment, an *Analysis Layer* composed of domain-specific Intelligent Specialized Agents (ISAs) for fundamental, technical, news, and sectoral analysis, and an *Integration Reasoning Layer* that fuses cross-modal outputs into coherent insights. A final *Decision Layer* transforms these insights into actionable buy/hold/sell recommendations with embedded risk assessment. By combining hierarchical multi-agent orchestration [4], structured inter-agent communication [5], and depth–breadth reasoning strategies [8], P1GPT advances beyond existing single-agent financial LLMs. Our contributions are threefold: (i) we propose the first end-to-end multi-agent LLM workflow tailored for multi-modal financial analysis, (ii) we design a structured reasoning and communication protocol to improve interpretability and reliability, and (iii) we demonstrate through backtesting that P1GPT achieves superior cumulative returns and risk-adjusted performance compared with rule-based and single-agent LLM baselines.

2 Related Work

2.1 Financial LLMs and Automated Assistants

Recent advances in domain-specific large language models (LLMs) have given rise to a series of financial LLMs and assistant systems tailored for market analysis. BloombergGPT [1] introduced a 50B-parameter model trained on extensive financial corpora, demonstrating substantial improvements across multiple financial NLP benchmarks. Similarly, FinGPT [2, 9, 10] provides an open-source ecosystem, later extended with Internet-scale data pipelines and instruction-tuning benchmarks, that supports benchmark evaluation and instruction tuning for financial applications, lowering the entry barrier for research and deployment. MarketSenseAI 2.0 [11] integrates retrieval-augmented generation with LLM agents to process earnings reports, news streams, and macroeconomic indicators, thereby enabling holistic stock analysis. Recent empirical studies further evaluated MarketSenseAI with GPT-4 in financial decision-making tasks, suggesting its potential while highlighting the challenge of consistently outperforming traditional benchmarks [12]. Beyond general analysis, FinLlama [6] leverages LLMs for financial sentiment classification to assist algorithmic trading. InvestLM [13], built on LLaMA-65B with domain-specific instruction tuning, achieves expert-level performance in investment reasoning tasks. More recent works such as Fin-R1 [14] highlight reinforcement learning–enhanced financial reasoning, while Agentar-Fin-R1 [15] extends this direction with domain-expert knowledge, trust-oriented data governance, and advanced multi-stage training strategies. In parallel, Bridging Language Models and Financial Analysis [16] explores the integration of textual, numerical, and tabular data within LLM pipelines. Collectively, these efforts highlight the growing trend of positioning LLMs as financial assistants, yet they remain largely single-agent in nature and often lack structured multi-agent workflows or scalable decision orchestration mechanisms.

2.2 LLMs as Autonomous Traders

Recent research has also explored the potential of large language models as autonomous trading agents that directly generate buy, hold, or sell decisions without relying on multi-agent orchestration. For instance, Can Large Language Models Trade? [3] investigates whether LLMs can act as traders in simulated financial markets, testing their ability to validate classical financial theories under market dynamics. Building on this line of work, FLAG-Trader [17] integrates LLM-based policy networks with gradient-based reinforcement learning, demonstrating improved sequential decision-making in trading environments. Other studies highlight the role of sentiment-driven strategies, where LLMs analyze news and social media signals to produce actionable trading recommendations, as in Sentiment Trading with Large Language Models [18]. Beyond direct trading signals, QuantAgent [19] employs a self-improving loop in which LLMs generate and refine alpha factors, showing promise for automated strategy discovery. Similarly, Evaluating LLMs in Financial Tasks [20] benchmarks LLMs on strategy code generation tasks, revealing both opportunities and limitations in translating natural language prompts into executable trading logic. More recently, Analysis of LLM Agent

Behavior in Experimental Asset Markets [21] examined how LLM traders behave in simulated bubbles and crashes, providing insights into emergent market dynamics. Additionally, MountainLion [22] demonstrates a multi-modal agent system in which each LLM operates as an autonomous trader, producing interpretable and adaptive strategies in volatile market conditions. Overall, these works illustrate the feasibility of LLMs functioning as standalone trading entities, yet they often lack mechanisms for structured collaboration and comprehensive multi-modal integration.

2.3 Multi-Agent Systems in Financial Applications

Recent studies have increasingly explored multi-agent LLM frameworks to capture the collaborative and organizational dynamics of financial decision-making. TradingAgents [4] models a virtual trading firm composed of analysts, researchers, traders, risk managers, and coordinators, demonstrating how role-based collaboration can enhance explainability and risk control. Similarly, the FinAgent framework, presented as A Multimodal Foundation Agent for Financial Trading [5], integrates tool augmentation, reflective mechanisms, and diversified agents to support multi-modal financial analysis. FinVision [23] advances this direction by orchestrating multiple agents for stock market prediction, combining technical, news, and reflective modules. Beyond application-driven designs, MASS [7] introduces large-scale multi-agent simulations—scaling up to 512 agents—to optimize portfolio construction and uncover scaling laws in investor modeling. Other works such as FinCon [24] mimic hierarchical financial organizations with verbal reinforcement mechanisms to improve decision robustness, while HedgeAgents [25] emphasizes risk-aware collaboration through hedging strategies. More recent proposals extend multi-agent reasoning to deeper financial contexts: an analytical stage-based multi-agent framework enhanced with expert knowledge [26] improves fundamental analysis, and Automate Strategy Finding with LLM in Quant Investment [27] explores risk-aware multi-agent systems for quantitative strategy discovery. From a broader theoretical perspective, the Dual Engines of Thoughts (DEoT) framework [8] introduces a depth-breadth integration mechanism for open-ended reasoning, offering a foundational perspective for layered multi-agent decision workflows. Together, these works illustrate the potential of multi-agent architectures to provide more scalable, interpretable, and realistic financial analysis compared with single-agent LLM approaches.

Unlike prior systems, P1GPT couples a structured, layer-wise workflow with standardized agent reports and a dedicated integration reasoning layer, enabling reproducible, auditable decisions and scalable multi-modal fusion across market regimes.

2.4 Comparison with Existing Frameworks

Several representative frameworks have recently explored the intersection of LLMs and financial analysis, yet each exhibits limitations that motivate the design of P1GPT. MarketSenseAI focuses on retrieval-augmented generation with a primarily single-agent setup, achieving solid text and data integration but lacking multi-agent coordination. TradingAgents models a virtual trading team with distinct roles such as analysts and risk managers, offering high interpretability but relying on free-form natural language exchanges that may reduce efficiency. FinAgent emphasizes tool use and multimodal inputs, yet its design remains largely experimental without a structured layered workflow. MASS investigates scalability with hundreds of agents, but sacrifices interpretability and practical deployment. FinVision combines technical and news-driven modules in a multi-agent setting, though it lacks rigorous integration and decision layers.

Framework	Architecture	Explainability	Practicality
MarketSenseAI	Single-agent + RAG; text/data/news focused	Medium (retrieval-centered)	Research-oriented, limited coordination
TradingAgents	Multi-agent, role-based (analyst, trader, risk)	High (clear roles)	Simulation focus, less deployable
FinAgent	Multi-agent with tool usage	Medium	Expandable, but experimental
MASS	Large-scale simulation (100+ agents)	Low	Studies scalability, weak interpretability
FinVision	Multi-agent (technical + news)	Medium	Academic focus, lacks integration/decision layers
P1GPT (ours)	Layered multi-agent workflow with structured communication	High (standardized, traceable reports)	Modular, extensible, deployment-ready

Table 1: Comparison of representative financial LLM frameworks with P1GPT.

Compared with these frameworks, **P1GPT** provides a layered multi-agent architecture that not only supports comprehensive multimodal data integration but also enforces structured communication and standardized reporting. This design bridges the gap between academic prototypes and deployable financial systems, positioning P1GPT as a practical foundation for intelligent trading workflows.

3 P1GPT System Overview

3.1 Overall System Architecture

P1GPT adopts a modular, layered architecture that mirrors the decision-making flow of a modern multi-agent financial analysis system. As illustrated in Figure 1, the framework is organized into five distinct layers, each responsible for a critical stage of the reasoning pipeline:

1. **Input Layer:** Parses user queries and multi-modal financial data through entity extraction, query understanding, and data preprocessing.
2. **Planning Layer:** Decomposes tasks, assigns them to appropriate Intelligent Specialized Agents (ISA), and maps inter-agent dependencies based on query intent and data types.
3. **Analysis Layer:** Executes task-specific analysis through domain-specialized agents (e.g., Fundamental ISA, News ISA), while managing agent coordination and scheduling.
4. **Integration Layer:** Aggregates analytical outputs, performs cross-agent reasoning, and synthesizes insights into coherent intermediate summaries and structured reports.
5. **Decision Layer:** Generates actionable investment signals, conducts risk assessment and strategy planning, and produces natural language summaries as the final system output.



Figure 1: P1GPT overall workflow. Left: **Input** and **Planning** transform user queries and raw multi-modal data into an executable plan. Center: the **Analysis Layer** hosts specialized agents—Controller, Fundamental ISA, Tech ISA, Chip (Semiconductor) ISA, News ISA—and supporting modules (Search, Revenue, Trend Analyzer, Recommender). Right: the **Integration Layer** consolidates structured reports and forwards them to the **Decision Layer**, which outputs an *Action* with an explanation. Arrows denote data/control flow and standardized report interfaces between layers.

This structured architecture enables modular orchestration, parallel agent execution, and complete workflow traceability. Each layer is independently extensible, ensuring that the system remains adaptable to evolving financial contexts. The detailed workflow of these layers is presented in Section 4.

3.2 Data Sources and Collection Pipeline

P1GPT is built to analyze multi-modal financial data, integrating a diverse set of information streams to enhance decision accuracy and robustness. The collection pipeline comprises the following major data sources:

- **News and Financial Media:** Real-time news articles, press releases, and macroeconomic reports are collected from global financial news providers (e.g., Bloomberg, Reuters, Yahoo Finance) via API or RSS feeds. These sources provide timely updates on market events, company actions, and regulatory changes.
- **Social Media and Forums:** Posts and discussions from platforms such as X (formerly Twitter), Reddit (e.g., r/StockMarket, r/wallstreetbets), and local forums are scraped using dedicated APIs or third-party aggregators. Sentiment and engagement metrics are extracted to assess retail investor mood and crowd reactions.

- **Technical Market Data:** Historical and real-time price series, trading volumes, and a suite of technical indicators (e.g., MACD, RSI, SMA) are ingested from financial market data providers. This data forms the basis for technical trend analysis and quantitative modeling.
- **Fundamental Data:** Company financial statements, earnings reports, balance sheets, analyst estimates, and insider transactions are obtained from open databases (e.g., Yahoo Finance, Finnhub, EDGAR) and processed for fundamental valuation and risk assessment.

The pipeline automates the entire ingestion process: raw data is fetched at regular intervals, normalized into a unified schema, and stored in a centralized data repository. Data pre-processing includes entity recognition, timestamp alignment, duplicate removal, and language normalization. For textual data, natural language processing (NLP) models are used to extract relevant entities, classify sentiment, and generate summaries. All processed and standardized data is then made available to the relevant agents via a shared access layer, ensuring that each agent can query the latest information for its analytical tasks.

This comprehensive and automated pipeline enables P1GPT to maintain an up-to-date and holistic view of the financial environment, supporting timely and well-informed multi-agent reasoning.

3.3 Core Modules and Agent Roles

P1GPT employs a modular multi-agent design, inspired by the structure of real-world financial analysis teams and the architecture of the Summarizer system. Each agent in the system is responsible for a specialized analytical task, contributing to a coordinated and explainable decision-making pipeline. These agents are organized into two categories: *Intelligent Specialized Agents (ISA)* and *Supporting Agents*.

3.3.1 Intelligent Specialized Agents (ISA)

The ISA family consists of domain-specialized agents designed to handle core financial analysis functions. Each ISA processes a particular modality or market perspective using LLMs and financial tools:

- **Controller ISA:** Serves as the central coordinator of the agent workflow. It assigns tasks, manages execution order, and integrates analytical outputs to ensure coherent reasoning and final synthesis.
- **Fundamental Analysis ISA:** Analyzes company fundamentals, including earnings reports, balance sheets, valuation ratios, and analyst expectations. It provides long-term investment insights based on corporate financial health.
- **Technical Analysis ISA:** Interprets market signals from historical price and volume data by applying technical indicators such as MACD, RSI, and moving averages. It supports short- to medium-term trading strategies.
- **Semiconductor ISA:** Focuses on the semiconductor and supply chain domain, aggregating and interpreting industry news, market trends, and innovation signals to assess sector-specific opportunities and risks.
- **News Intelligence ISA:** Aggregates financial news, regulatory updates, and press releases. It applies NLP techniques to summarize content, extract sentiment, and identify events that may influence market dynamics.

3.3.2 Supporting Agents

These agents enhance the capabilities of the core ISA modules by providing external retrieval, forecasting, and decision support functionalities:

- **External Search Agent:** Performs real-time information retrieval from external knowledge bases and APIs (e.g., Perplexity AI) to supplement ISA analysis with up-to-date facts and context.
- **Revenue Forecasting Agent:** Utilizes statistical models and historical financial data to predict future revenue or earnings figures for target companies, aiding in valuation and scenario testing.
- **Investment Recommendation Agent:** Synthesizes the outputs of all analytical agents to formulate actionable investment decisions (e.g., buy/sell/hold), along with confidence scores and supporting rationales.
- **Market Trend Analysis Agent:** Tracks broader market indices, trading volumes, and price momentum to identify macroeconomic trends, sectoral rotations, and long-term patterns.

All agents operate as independent modules with defined input/output interfaces and communicate through a structured protocol. Their outputs are formatted into standardized reports, which are aggregated and evaluated by the Controller ISA and the decision-making layer. This architecture ensures flexibility, scalability, and robust multi-modal reasoning across complex financial environments.

3.4 Integration Strategy and Data Fusion

To maximize the value of multi-modal information, P1GPT implements a structured integration strategy that systematically fuses the outputs of all specialized agents. Each analytical agent produces a standardized report summarizing its findings, which are submitted to the Overall ISA Agent for aggregation and cross-comparison.

The integration process involves both rule-based and LLM-driven fusion techniques. Quantitative results from technical and fundamental analyses are normalized and combined with qualitative insights from news and sentiment agents. A centralized decision module evaluates the consistency and reliability of each agent's input, assigning dynamic weights or confidence scores according to the context and recent agent performance.

Where discrepancies or conflicts exist between agent assessments, the system may trigger additional analysis (e.g., calling the Perplexity Search Agent for external verification or requesting deeper analysis from a relevant agent). This iterative information refinement ensures that all available perspectives are considered before decision-making.

The final, integrated view supports robust and explainable trading decisions by synthesizing numerical scores, textual rationales, and trend predictions into a single actionable recommendation. This data fusion approach enhances both the breadth and depth of system reasoning, enabling the P1GPT framework to adapt flexibly to complex and rapidly changing market environments.

4 Layered Workflow and Communication

The P1GPT system is designed around a layered architecture that facilitates modularity, transparency, and collaboration among agents. From the initial user query to the final investment recommendation, each layer transforms, enriches, and routes information to the next stage of the pipeline. This section details the complete operational flow across the five core layers.

4.1 Input Layer

The Input Layer is responsible for transforming raw inputs—both user queries and external data streams—into structured representations suitable for downstream agent processing. It ensures that all subsequent reasoning is grounded in clean, interpretable, and context-aware information. Its core objectives include:

- **User Query Understanding** Natural language instructions are parsed to extract intent, target entities (e.g., tickers, indicators), and temporal scope using domain-adapted NLP models.
- **Entity and Feature Extraction** Financial entities such as companies, metrics, and events are identified and tagged from both queries and raw textual data (e.g., news articles, forum posts).
- **Data Preprocessing and Harmonization** Structured (e.g., price data), semi-structured (e.g., filings), and unstructured sources (e.g., social media) are normalized into a unified format, with noise removed and timestamps aligned.
- **Input Structuring for Planning Layer** Parsed queries and preprocessed data are encapsulated into standardized input objects annotated with semantic metadata, enabling efficient consumption by the Planning Layer and Intelligent Specialized Agents.

By serving as a semantic gateway for both user intent and real-world market signals, the Input Layer ensures accuracy, clarity, and consistency throughout the P1GPT analytical workflow.

4.2 Planning Layer

The Planning Layer translates structured inputs from the Input Layer into an executable multi-agent workflow. It determines the analytical scope, decomposes complex tasks, and orchestrates agent collaboration to ensure comprehensive and efficient coverage of the problem space. Its key functions include:

- **Task Decomposition** Breaks down high-level analysis objectives into granular subtasks (e.g., fundamental analysis, sentiment extraction, trend detection) that can be executed independently or in sequence.
- **Agent Matching** Maps each subtask to the most suitable Intelligent Specialized Agent (ISA) or supporting agent based on domain expertise, required data modality, and computational capabilities.
- **Dependency Mapping** Establishes inter-task dependencies and execution order, enabling the system to run agents in parallel when possible while enforcing sequential processing where data flow dictates.

- **Dynamic Workflow Adaptation** Adjusts task allocation in real time if new data becomes available, user intent changes, or intermediate results suggest alternative analytical paths.

Through precise decomposition, optimal agent assignment, and dependency-aware scheduling, the Planning Layer ensures that every query triggers a coordinated, resource-efficient multi-agent reasoning process.

4.3 Analysis Layer

The Analysis Layer is the operational core of P1GPT, where domain-specialized reasoning is performed by both Intelligent Specialized Agents (ISAs) and supporting agents. Upon receiving the execution plan from the Planning Layer, each assigned agent processes its designated data segment using domain-specific methodologies and models.

ISAs cover critical analytical domains such as fundamentals, technical indicators, sector-specific intelligence (e.g., semiconductors), and market news sentiment. Each agent operates as an autonomous processing unit, equipped with tailored analytical pipelines—ranging from financial ratio evaluation and time-series modeling to natural language processing for event extraction. Supporting agents, such as external search and revenue forecasting modules, are invoked when additional contextual knowledge or predictive insights are required.

Inter-agent coordination is enabled through a shared communication protocol. Agents can request clarification from peers, exchange intermediate findings, and collaboratively resolve conflicting interpretations. This coordination mechanism is particularly important when dealing with multi-modal inputs, where insights from different domains must be reconciled before integration.

By combining independent domain expertise with controlled information exchange, the Analysis Layer ensures that the overall reasoning process remains both comprehensive and internally consistent. Its outputs—structured analytical reports annotated with confidence scores and explanatory rationales—form the foundation for the subsequent integration and decision-making stages.

4.4 Integration Layer

The Integration Reasoning Layer consolidates the heterogeneous outputs generated by the Analysis Layer into a unified, coherent analytical perspective. This stage is responsible for aligning, reconciling, and synthesizing results across multiple modalities and domains to ensure that subsequent decision-making is both informed and explainable.

Integration begins with the collection of structured reports from all participating agents. These reports contain standardized elements such as key findings, supporting evidence, confidence scores, and domain-specific rationale. An alignment process maps the outputs to shared reference entities—such as stock tickers, timeframes, or economic indicators—allowing cross-domain comparisons and eliminating redundancies.

Reasoning within this layer is performed through a combination of deterministic rules and LLM-driven synthesis. Rule-based logic is employed to enforce domain priorities and resolve conflicts (e.g., prioritizing fundamental analysis over short-term sentiment in long-horizon strategies), while large language models are used to produce higher-level interpretations that integrate quantitative metrics with qualitative insights.

When conflicting or uncertain conclusions emerge, the layer may trigger iterative feedback loops with the Analysis Layer, requesting clarification, additional computations, or external validation. This ensures that discrepancies are addressed before final recommendations are produced.

By merging structured numerical signals with narrative reasoning, the Integration Reasoning Layer provides a complete and context-aware assessment of the financial environment, serving as the critical bridge between raw analysis and actionable decision-making.

4.5 Action Decision Layer

The Action Decision Layer transforms the synthesized insights from the Integration Reasoning Layer into concrete, actionable outputs for end-users or automated trading systems. Its primary objective is to bridge the gap between analytical assessment and operational execution while maintaining transparency and traceability.

This layer is driven by specialized decision-making agents, most notably the Investment Recommendation Agent and the Market Trend Analysis Agent. The former evaluates the integrated signals against predefined investment strategies, risk tolerance parameters, and market conditions to produce explicit trade recommendations—such as Buy, Sell, or Hold—accompanied by confidence scores and justifications. The latter focuses on identifying and validating mid- to long-term market trends, providing contextual guidance that can influence position sizing and portfolio adjustments.

Decisions are generated through a hybrid process: quantitative thresholds and risk metrics ensure that recommendations meet objective performance and safety criteria, while LLM-based reasoning modules articulate the rationale in natural language, enabling human users to understand and audit the underlying logic.

Where execution is automated, this layer interfaces directly with trading platforms or portfolio management systems through secure APIs, ensuring timely and accurate order placement. In human-in-the-loop scenarios, the recommendations are presented via an interactive dashboard, allowing analysts or traders to review, adjust, or override system-generated actions.

By tightly coupling analytical rigor with operational delivery, the Action Decision Layer ensures that P1GPT’s outputs are not only theoretically sound but also practically implementable in dynamic financial environments.

5 Experiments

5.1 Simulation Setup

To evaluate the performance of **P1GPT** in realistic financial trading scenarios, we conduct comprehensive backtesting simulations using multi-modal historical data. The simulation focuses on three representative U.S. equities—**AAPL**, **GOOGL**, and **TSLA**—which collectively span major innovation-driven sectors: consumer technology, digital platforms, and electric vehicles. These companies were selected because they represent distinct yet interconnected domains that are highly sensitive to macroeconomic shifts, supply chain policies, and investor sentiment. Analyzing these stocks allows the evaluation of P1GPT’s adaptability across heterogeneous market structures and volatility regimes. The evaluation period spans from **February 1, 2025** to **September 30, 2025**, a phase marked by notable post-election market uncertainty and renewed trade policy dynamics.

In particular, the reinstatement of import tariffs under the Trump administration in early 2025 triggered sector-specific volatility—especially within technology and semiconductor supply chains—providing a realistic stress test for the system’s reasoning-driven decision process. This window also includes several quarterly earnings releases and sentiment reversals, enabling assessment of how P1GPT responds to rapidly evolving multi-modal financial contexts.

For each trading day, the system receives a structured prompt that includes the current market context and a target stock ticker. All data are preprocessed through the collection and normalization pipeline described in Section 3.2 to ensure consistency and temporal alignment. The simulation strictly avoids lookahead bias—agents can only access information available prior to the decision timestamp.

Baselines. To benchmark the performance of P1GPT, we compare it against five established strategies following the TradingAgents evaluation protocol:

- **Buy & Hold (B&H):** Passive investment throughout the simulation window.
- **MACD:** Moving Average Convergence Divergence crossover strategy.
- **KDJ + RSI:** Momentum-based signals combining stochastic oscillator and RSI thresholds to detect overbought/oversold conditions.
- **ZMR:** Zero-Mean Reversion strategy using deviations from an equilibrium price as trading triggers.
- **SMA:** Simple moving-average cross strategy (10-day vs. 20-day windows).

These baselines provide diverse perspectives on market behavior—ranging from passive exposure to active momentum and reversion-based trading—allowing a fair evaluation of P1GPT’s reasoning-driven decision process.

Evaluation Metrics. Four standard financial metrics are used to evaluate the profitability, risk, and stability of each strategy:

- **Cumulative Return (CR):**

$$CR = \frac{V_{\text{end}} - V_{\text{start}}}{V_{\text{start}}} \times 100\%$$

where V_{start} and V_{end} denote initial and final portfolio values, respectively.

- **Annualized Return (AR):**

$$AR = \left(\frac{V_{\text{end}}}{V_{\text{start}}} \right)^{1/N} - 1$$

where N represents the number of years in the simulation period.

- **Sharpe Ratio (SR):**

$$SR = \frac{\bar{R} - R_f}{\sigma}$$

where \bar{R} is the average daily return, R_f the risk-free rate (approximated by 3-month U.S. Treasury yield), and σ the standard deviation of daily returns.

- **Maximum Drawdown (MDD):**

$$MDD = \max_{t \in [0, T]} \left(\frac{\text{Peak}_t - \text{Trough}_t}{\text{Peak}_t} \right) \times 100\%$$

which represents the largest observed loss from a peak to a trough within the simulation window.

5.2 Backtesting Simulation

For each trading day t , P1GPT receives a structured query of the form: "Given today's market conditions, should I buy, sell, or hold [TICKER]?" This query is processed through the system's hierarchical reasoning pipeline, activating specialized agents in the Fundamental, Technical, News, and Sentiment domains. Each agent contributes domain-specific insights that are synthesized within the Integration and Decision Layers to produce a final investment signal.

The system outputs a single trading decision—**Buy**, **Sell** or **Hold**—accompanied by a concise natural-language explanation summarizing the reasoning process. No numerical confidence score is produced; instead, interpretability is achieved through the transparent textual rationale and structured inter-agent reports.

A simple execution policy governs portfolio operations:

- **Buy:** Enter a position at the same day's closing price.
- **Sell:** Exit any currently held position.
- **Hold:** Maintain the existing position without modification.

The backtesting assumes no transaction costs, no leverage, and a single-stock position constraint per run. Daily portfolio values are aggregated to compute CR, AR, SR, and MDD across the test horizon. The resulting statistics enable an objective comparison between P1GPT and baseline strategies in terms of profitability, volatility control, and drawdown resilience.

6 Results, Analysis, and Discussion

6.1 Quantitative Performance Evaluation

The quantitative outcomes are summarized in Table 2, accompanied by Figures 2–5, which illustrate P1GPT's equity growth trajectories on AAPL, GOOGL, and TSLA. Together, they reveal that **P1GPT** maintains a well-calibrated balance between profitability and risk, consistently outperforming traditional rule-based strategies across diverse market environments.

Table 2: Performance comparison across all methods using four evaluation metrics. Results highlighted in green represent the best-performing statistic (higher is better for CR/AR/SR; lower is better for MDD).

Categories	Models	AAPL				GOOGL				TSLA			
		CR%↑	AR%↑	SR↑	MDD%↓	CR%↑	AR%↑	SR↑	MDD%↓	CR%↑	AR%↑	SR↑	MDD%↓
Market	B&H	2.66	4.07	0.56	7.33	4.19	6.42	1.05	6.41	6.10	9.41	0.52	16.89
Rule-based	MACD	-1.87	-2.83	-0.59	4.79	2.67	4.09	0.83	4.11	7.36	11.38	0.88	10.69
	KDJ+RSI	2.92	4.46	1.24	1.78	0.98	1.49	0.60	1.28	8.65	13.43	1.13	6.37
	ZMR	2.50	3.83	0.64	5.46	-0.62	-0.94	-0.22	4.69	3.61	5.53	0.41	15.60
	SMA	-3.73	-5.60	-1.03	7.57	8.23	12.76	2.90	1.32	6.31	9.73	0.68	11.43
Ours	P1GPT	16.16	25.53	3.38	2.02	31.64	51.78	2.71	7.08	22.79	36.57	2.31	6.50

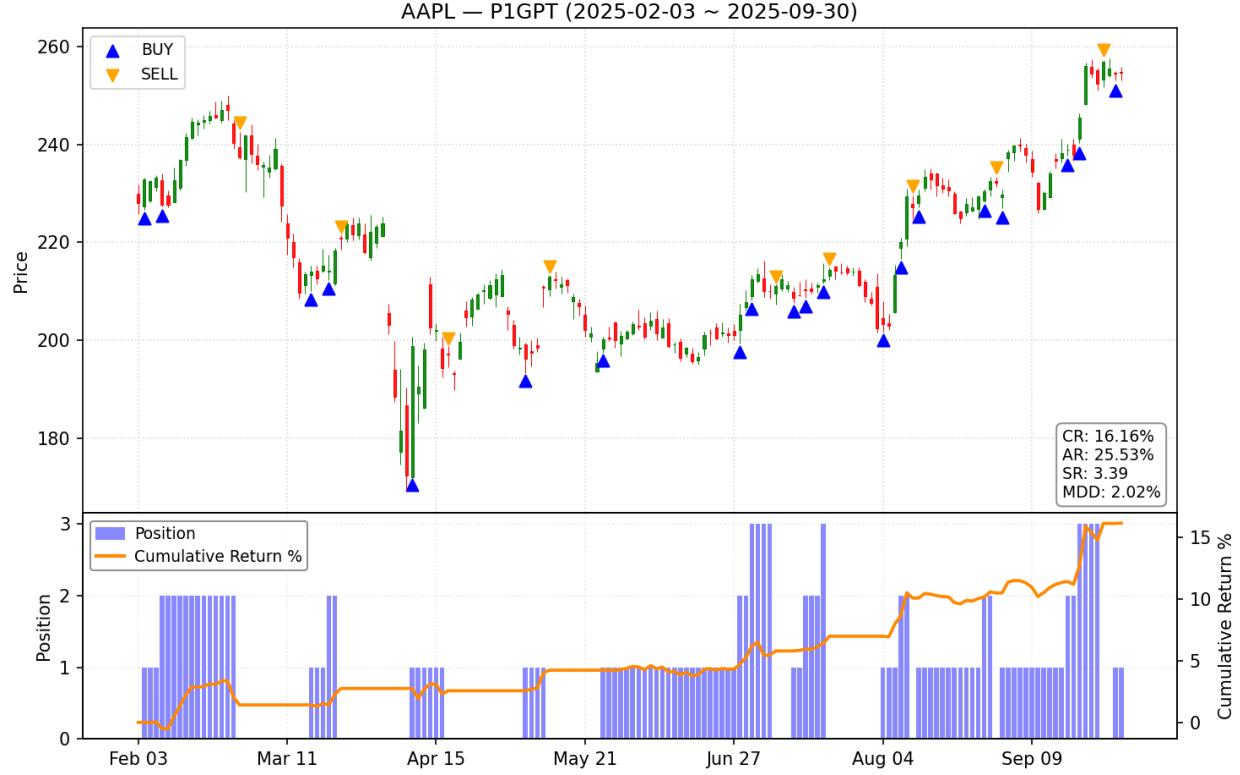


Figure 2: Trading trajectory of P1GPT on AAPL (Feb. 3–Sep. 30, 2025), showing model-generated buy (blue) and sell (orange) signals, along with position and cumulative return evolution.

6.1.1 Cumulative and Annual Returns

P1GPT achieves the highest profitability across all tested assets, with cumulative returns exceeding 16% on AAPL, 31% on GOOGL, and 22% on TSLA. Its annualized returns surpass the best rule-based baselines by over 20 percentage points on average. This improvement stems from its hierarchical reasoning framework, which integrates multi-modal signals—including earnings trends, technical momentum, and sentiment shifts—allowing adaptive repositioning during changing market phases. Unlike static threshold systems such as SMA or MACD, P1GPT’s dynamic synthesis produces smoother return curves and sustained compounding over time.

6.1.2 Sharpe Ratio

P1GPT maintains strong risk-adjusted performance, achieving Sharpe ratios of roughly 3.4 (AAPL), 2.7 (GOOGL), and 2.3 (TSLA). Although SMA slightly exceeds it on GOOGL, P1GPT exhibits greater cross-asset consistency, reflecting a deliberate risk-return calibration rather than volatility suppression. By tolerating measured fluctuations when they contribute to higher expected payoff, the system converts information advantage into reward without relying on excessive leverage or reactive timing.

6.1.3 Maximum Drawdown

The framework’s maximum drawdowns remain controlled—around 2% for AAPL and below 7% for the other assets—demonstrating resilience under market stress. While certain conservative baselines achieve marginally lower drawdowns (e.g., KDJ+RSI on GOOGL at 1.3%), they do so at the cost of diminished returns. P1GPT’s behavior reflects a balanced posture: accepting limited short-term losses to capture larger long-term gains, achieving stability through informed adaptability rather than rigid constraint.

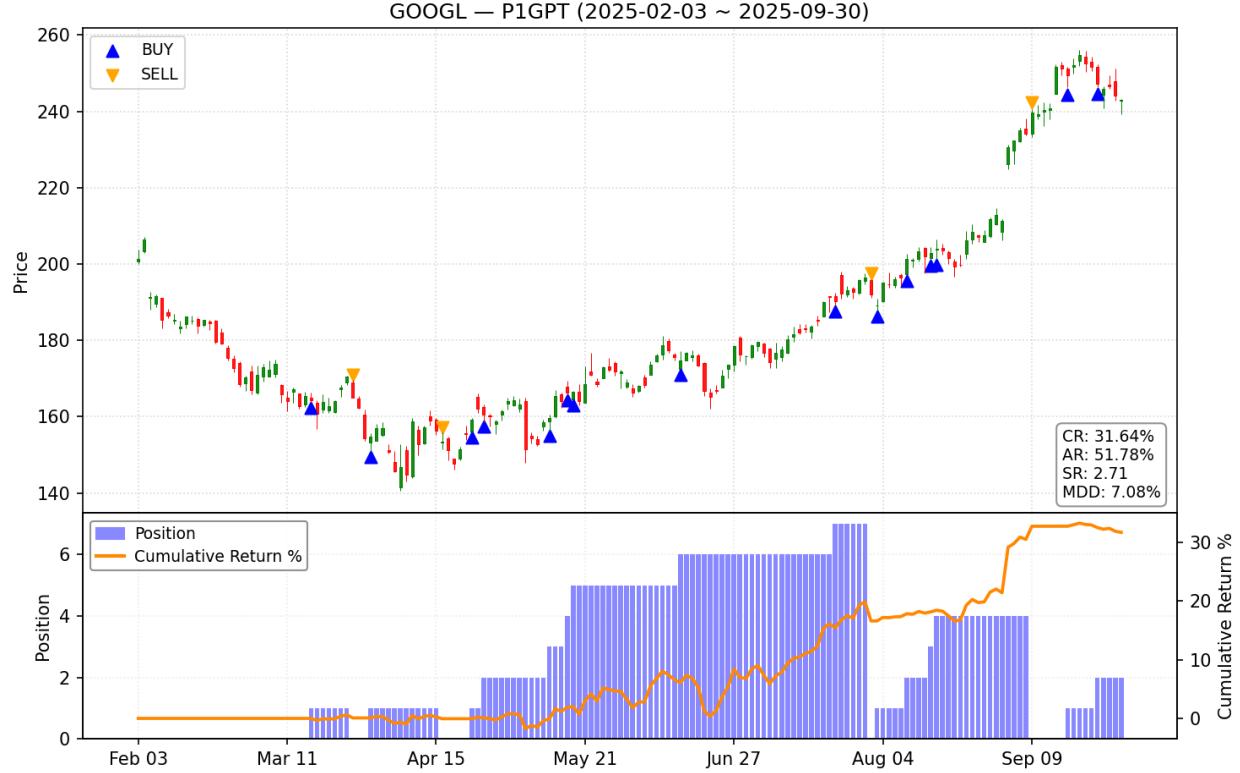


Figure 3: Trading trajectory of P1GPT on GOOGL (Feb. 3–Sep. 30, 2025), showing model-generated buy (blue) and sell (orange) signals, along with position and cumulative return evolution.

6.1.4 Cross-Asset Robustness

Across the three sectors—consumer technology, digital services, and electric vehicles—P1GPT consistently outperforms baseline strategies in cumulative and annualized returns while maintaining Sharpe ratios above 2.0 and drawdowns within moderate bounds. This consistency underscores its generalization capability: the reasoning-driven, multi-agent architecture transfers effectively across market regimes, delivering interpretable and risk-aware performance rather than over-optimized results specific to one asset or indicator.

6.2 Qualitative Behavioral Analysis

6.2.1 Overview of Behavioral Patterns

Figures 2–4 visualize the trading trajectories of P1GPT across three representative equities. Across all assets, the system exhibits a consistent behavioral pattern characterized by *adaptive re-entry*, *risk-aware exit*, and *volatility-conscious exposure control*. Rather than performing frequent short-term trades, P1GPT demonstrates a structured rhythm of position building and unwinding that reflects reasoning-driven judgment rather than reactive signal following.

For **AAPL**, the system repeatedly re-enters positions shortly after localized drawdowns and avoids holding through extended downtrends. This behavior suggests a disciplined recovery strategy—reallocating exposure once technical and fundamental conditions realign—leading to steady cumulative gains with limited volatility. In contrast to rule-based methods that rely on fixed thresholds, P1GPT’s actions adapt dynamically to changing market regimes, as seen in its avoidance of early entries during the March downturn and its timely re-engagement around July–August when momentum stabilized.

For **GOOGL**, P1GPT’s trade sequence reveals a gradual accumulation phase during the mid-year recovery, followed by profit-taking near short-term peaks. The position histogram indicates that the system increased exposure as confidence signals strengthened across multiple modalities (e.g., fundamental growth, sentiment improvement), aligning with the broader uptrend. Although its drawdown (**7.08%**) was higher than that of AAPL, the cumulative return (**31.64%**) shows that the model consciously accepted controlled risk in exchange for higher opportunity capture.

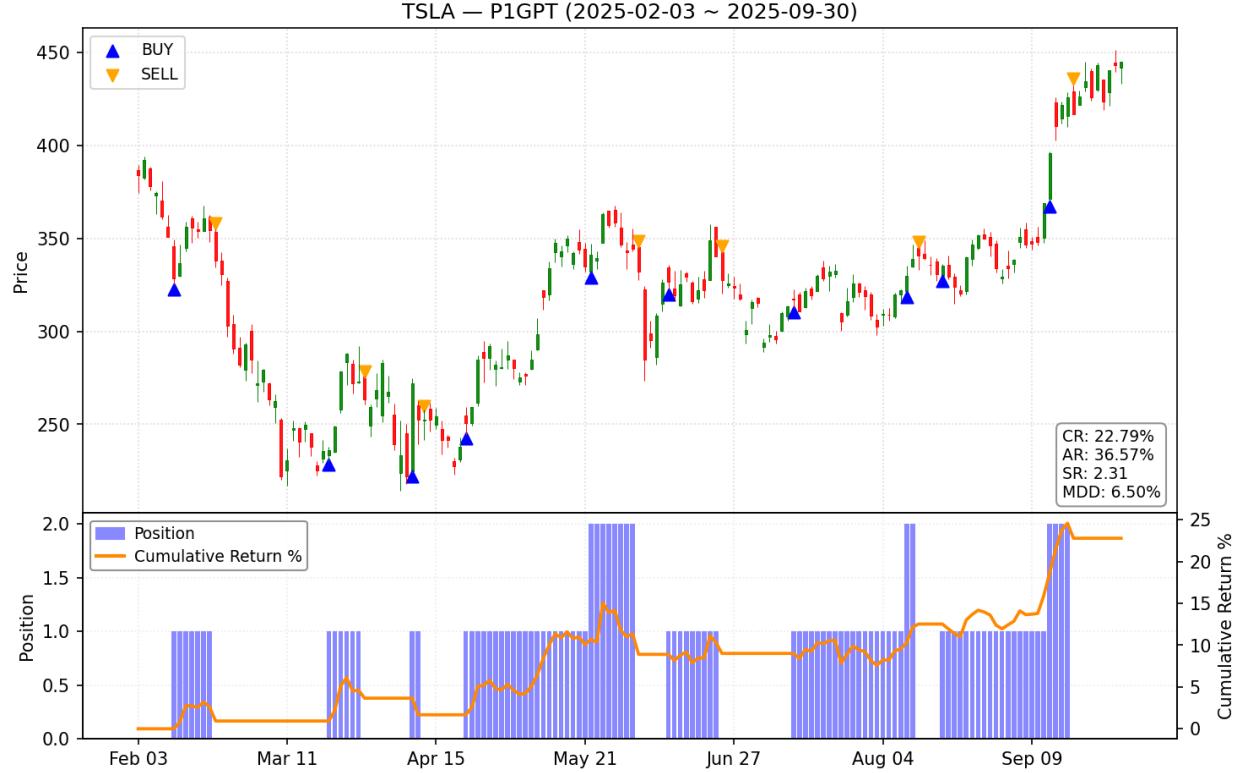


Figure 4: Trading trajectory of P1GPT on TSLA (Feb. 3–Sep. 30, 2025), showing model-generated buy (blue) and sell (orange) signals, along with position and cumulative return evolution.

On **TSLA**, the agent demonstrates cautious entry behavior during the volatile early period, favoring shorter holding intervals and early exits amid downward pressure. As sentiment and technical indicators reversed around mid-year, P1GPT transitioned toward longer holding durations and a pronounced upward exposure in late Q3. This behavioral shift illustrates its capacity for *contextual reasoning*: the model interprets fundamental recovery and sentiment reversal as sufficient justification for re-entry, achieving efficient timing without overtrading.

Taken together, these behavioral patterns illustrate P1GPT’s explainable investment logic: it internalizes risk–reward trade-offs through cross-agent consensus, forming a decision process that combines macro context, sector-specific trends, and short-term signals. Such reasoning-based adaptivity explains why the model achieves consistent profitability under heterogeneous market conditions while maintaining interpretable decision transparency.

6.2.2 Case-Level Event Analysis

To further illustrate P1GPT’s reasoning-driven decision process, we analyze two consecutive investment reports generated during March–April 2025 for **AAPL**. These cases demonstrate how the system synthesizes technical, fundamental, and news-based information to adapt its stance from a defensive *Sell* position to a confident *Buy* signal as market context evolves.

As shown in Figure 6, P1GPT initially adopted a cautious outlook on March 24, 2025. The **Technical ISA** detected weak short-term support near \$220, with RSI below 35 and price action well beneath the 20-day moving average, signaling persistent downward pressure. The **Fundamental ISA** noted an elevated valuation (trailing P/E 37.6) and high leverage (debt-to-equity 154.5), while the **News ISA** emphasized geopolitical uncertainty, cost risks from supply-chain relocation, and broader concerns over a possible AI valuation bubble. The consensus among agents concluded that AAPL was in a technically fragile phase with stretched valuation and fading momentum, leading to a *Sell or reduce holdings on strength* recommendation. This reasoning aligned with the prevailing macro backdrop at the time—renewed tariff tensions and cautious investor sentiment toward overvalued technology stocks.

However, as shown in Figure 7, the system reversed its position on April 10, 2025 as new information shifted the macro and technical landscape. The **Technical ISA** observed price stabilization near the \$183–185 support zone and RSI levels

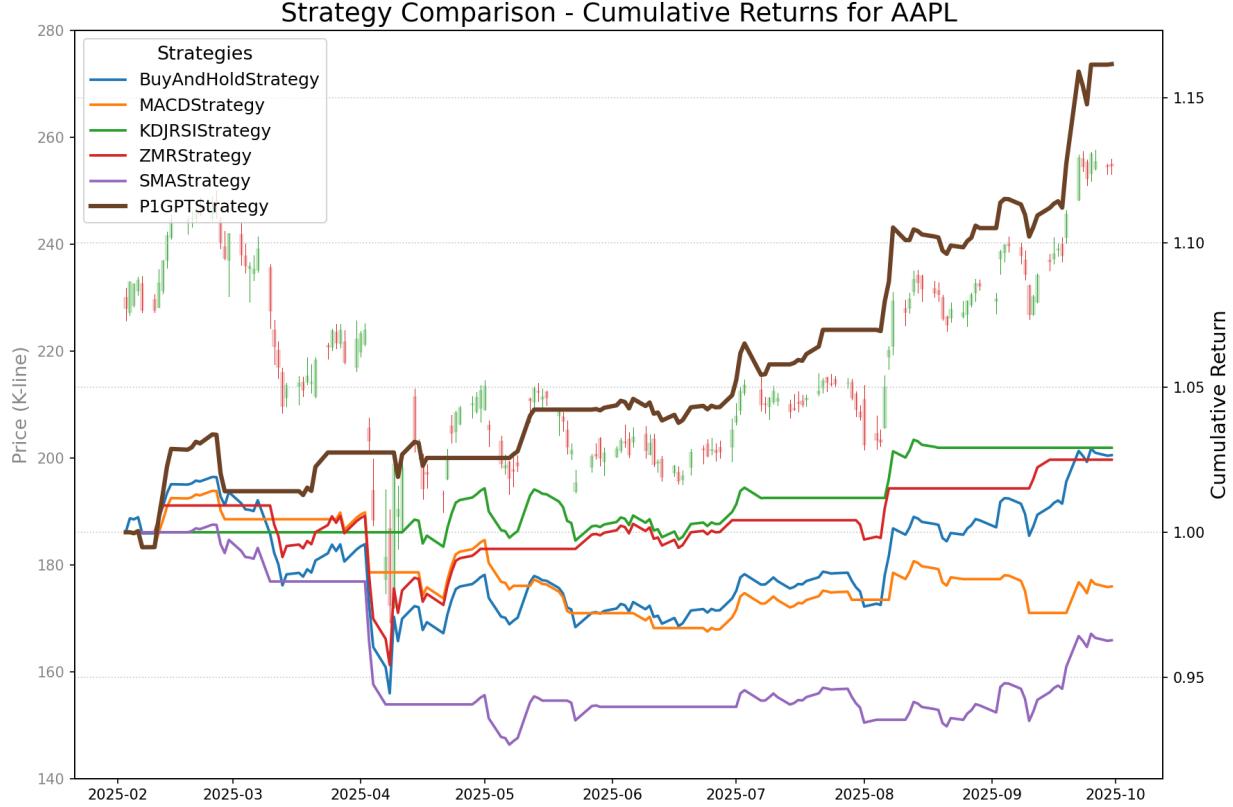


Figure 5: Cumulative Returns on AAPL using P1GPT. The figure shows the performance comparison of our model against baseline approaches for Apple Inc. stock analysis.

approaching oversold territory (~37), indicating the potential for a rebound. Simultaneously, the **News ISA** reported positive developments, including a meeting between China’s Industry Minister and Tim Cook and expectations of tariff relief under the Trump administration—signals of improving policy tone and investor sentiment. The **Fundamental ISA** confirmed Apple’s continued profitability (operating margin 29.9%) and robust free cash flow (\$94.9 B), supporting valuation resilience. Integrating these updates, the agents collectively upgraded the outlook to *Buy (accumulate on dips)*, anticipating a short-term recovery toward the \$210 resistance level. This transition from caution to accumulation exemplifies P1GPT’s ability to perform contextual reasoning—reassessing multi-modal inputs dynamically rather than following static momentum cues.

These paired cases highlight the explainable nature of P1GPT’s reasoning framework. The system’s decisions are not reactive to isolated indicators but emerge from the consensus of specialized agents evaluating technical exhaustion, valuation shifts, policy context, and sentiment direction in unison. The causal progression from *Sell* to *Buy* demonstrates reasoning coherence and adaptability under uncertainty, illustrating how P1GPT’s multi-agent architecture translates real-world evidence into interpretable and strategically consistent actions.

6.3 Discussion

Our findings indicate that P1GPT’s edge derives from mechanism design rather than indicator chasing: a modular division of labor, standardized inter-agent reporting, and a synthesis step that arbitrates evidence across modalities. This design yields an operational principle of *contextual exposure*: the system accepts risk only when multi-modal signals are aligned, and scales back when they conflict. Behaviorally, the agent exhibits stable patterns—measured re-entry after drawdowns and disciplined exits around regime shifts—that are *traceable* to agent rationales rather than emergent from opaque heuristics. Hence, qualitative transparency complements quantitative gains and supports auditability. Limitations remain. We abstract away transaction costs and portfolio interactions, and evaluate within a single market window. Extending to multi-asset allocation, realistic execution frictions, and formal causal tracing from agent outputs to actions are necessary steps to assess robustness and scalability.

Investment Report: AAPL (Apple Inc.)

Date: March 24, 2025

Technical Analysis

- **Current Price:** \$220.73, hovering near the recent low of \$218.58, indicating weak short-term support.
- **Relative Strength Index (RSI):** 34.7 — slightly oversold, yet lacking a clear rebound signal.
- **20-Day Simple Moving Average (SMA20):** \$226.71 — price remains well below this level, confirming a short-term downtrend.
- **Trading Volume:** 44,299,500 — elevated volatility without strong buying momentum.

Fundamental Analysis

- **Market Cap:** \$3.68 trillion — stable in scale, but valuation remains elevated.
- **Trailing P/E Ratio:** 37.6 — reflects optimistic expectations but raises overvaluation concerns.
- **Revenue Growth:** 9.6%, signaling a slowdown in momentum.
- **Debt-to-Equity Ratio:** 154.5, showing high leverage risk.
- **Free Cash Flow:** \$94.87 billion — solid liquidity, but not sufficient to offset valuation pressure.

News Overview

1. **Supply Chain Relocation to Vietnam:** Indicates diversification efforts but introduces near-term cost and execution risks.
2. **Market Volatility in Tech Sector:** Investor sentiment toward high-valuation tech stocks remains cautious.
3. **AI Valuation Bubble:** Analysts warn that AI-related stocks are overpriced, increasing correction risk.
4. **China Market Updates:** While iPhone Air preorders and eSIM approval are positive, weak consumer confidence limits upside potential.

Summary

Technically, AAPL remains in a weak phase below major resistance. Fundamentally, high valuation and leverage levels raise caution. Although news regarding supply-chain diversification and product launches provides long-term potential, short-term momentum is fading. The balance of evidence suggests limited rebound potential and growing downside pressure.

Risks

- **Overvaluation Risk:** Elevated P/E and P/B ratios make the stock vulnerable to sharp corrections.
- **Geopolitical and Supply Chain Risk:** Ongoing U.S.-China tensions may impact margins and production efficiency.
- **Sentiment Volatility:** A sector-wide pullback in AI or tech names could further drag AAPL lower.

Investment Conclusion

Given the **weak technical structure, stretched valuation, and unsettled market sentiment**, it is advisable to **Sell or reduce holdings on strength** and wait for a more favorable technical reversal or market sentiment recovery before reentry.

Figure 6: Case 1 — AAPL Investment Report (March 24, 2025). P1GPT’s agents collectively issued a *Sell* recommendation as technical weakness, elevated valuation, and negative news sentiment converged. The reasoning emphasized overvaluation risks, geopolitical uncertainty, and lack of short-term momentum.

7 Conclusion

We presented **P1GPT**, a layered multi-agent LLM workflow for multi-modal financial analysis that turns heterogeneous evidence into auditable trading decisions. The framework emphasizes modular specialization, structured communication, and integration-time reasoning—a combination that delivers interpretable decisions with competitive risk–return characteristics. Future work will focus on deployment realism (cost-aware execution, live paper-trading), portfolio-level extensions (multi-asset, hedging, and dynamic sizing), and deeper transparency (agent-weight adaptation and causal provenance of decisions). We view reasoning-centered, modular architectures as a promising direction for trustworthy financial AI.

Investment Report: AAPL (Apple Inc.)

Date: April 10, 2025

Technical Analysis

- **Price Action:** AAPL opened at \$189.07, hit a high of \$194.78, a low of \$183.00, and closed at \$190.42 — reflecting elevated volatility near recent lows.
- **Volume:** 121.88M shares traded, indicating strong participation and potential accumulation at lower levels.
- **SMA20:** \$209.97 — the stock remains below its short-term moving average, yet the gap is narrowing. **RSI (14):** 37.14 — approaching the oversold zone, suggesting potential for a rebound. **Support Zone:** The \$183–185 range has acted as a **key support area**, forming multiple lower wicks — a potential base for short-term recovery.

Fundamental Analysis

- **Market Cap:** \$3.68 trillion — maintaining Apple's leadership as the world's largest tech company.
- **Valuation:** **Trailing P/E:** 37.6 — slightly high but still within the range for mega-cap tech stocks. **Forward P/E:** 29.8 — reflects continued earnings confidence.
- **Leverage:** Debt-to-equity ratio of 154.5 — relatively high but well-covered by consistent free cash flow.

News Overview

1. **China's Industry Minister meets Tim Cook:** A positive diplomatic signal suggesting improved cross-border business sentiment.
2. **Trump Tariff Policy Outlook:** Market expectations point toward **possible tariff relief or delayed enforcement** on U.S. tech imports — a key **policy tailwind** supporting Apple's global supply chain stability.
3. **Supply Chain Diversification:** Apple's new Vietnam-based Home Hub and robotics projects strengthen its long-term risk mitigation strategy.

Summary

AAPL is trading near an established **technical support zone (USD 183–185)** with RSI levels approaching oversold. Combined with potential **policy support from Trump's tariff negotiations** and improving China relations, the setup favors a **short- to mid-term rebound**. While valuation remains elevated, Apple's strong fundamentals, robust cash flow, and diversified global operations provide resilience. If support holds, the stock could retest the \$210 level in the coming weeks.

Risks

- **Policy Uncertainty:** A breakdown in U.S.–China trade talks or renewed tariff escalation could limit upside.
- **Technical Risk:** A confirmed break below \$183 would invalidate the current rebound setup.
- **Competitive Pressure:** Strong competition from Samsung and Google in hardware and AI ecosystems.

Investment Conclusion

Given oversold technical conditions, improving macro sentiment, and the possibility of **Trump's tariff policy acting as a short-term stabilizer**, AAPL presents a **buying opportunity near the support zone**.

Recommended Action: BUY (Accumulate on dips within \$185–190 range)

Figure 7: Case 2 — AAPL Investment Report (April 10, 2025). Following market stabilization, P1GPT reversed its stance to a *Buy* signal, citing technical rebound near the \$185 support zone, positive policy outlook on U.S.–China tariffs, and improving sentiment indicators. This shift demonstrates the model's adaptive reasoning and multi-modal evidence integration.

References

- [1] 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.
- [2] Hongyang Yang, Xiao-Yang Liu, and Christina Dan Wang. Fingpt: Open-source financial large language models, 2023.
- [3] Alejandro Lopez-Lira. Can large language models trade? testing financial theories with llm agents in market simulations, 2025.
- [4] Yijia Xiao, Edward Sun, Di Luo, and Wei Wang. Tradingagents: Multi-agents llm financial trading framework, 2025.
- [5] Wentao Zhang, Lingxuan Zhao, Haochong Xia, Shuo Sun, Jiaze Sun, Molei Qin, Xinyi Li, Yuqing Zhao, Yilei Zhao, Xinyu Cai, Longtao Zheng, Xinrun Wang, and Bo An. A multimodal foundation agent for financial trading: Tool-augmented, diversified, and generalist. KDD '24, page 4314–4325, New York, NY, USA, 2024. Association for Computing Machinery.
- [6] Giorgos Iacovides, Thanos Konstantinidis, Mingxue Xu, and Danilo Mandic. Finllama: Llm-based financial sentiment analysis for algorithmic trading. In *Proceedings of the 5th ACM International Conference on AI in Finance, ICAIF '24*, page 134–141, New York, NY, USA, 2024. Association for Computing Machinery.
- [7] Taian Guo, Haiyang Shen, Jinsheng Huang, Zhengyang Mao, Junyu Luo, Zhuoru Chen, Xuhui Liu, Bingyu Xia, Luchen Liu, Yun Ma, and Ming Zhang. Mass: Multi-agent simulation scaling for portfolio construction, 2025.
- [8] Fei-Hsuan Yu, Yun-Cheng Chou, and Teng-Ruei Chen. Dual engines of thoughts: A depth-breadth integration framework for open-ended analysis, 2025.
- [9] Xiao-Yang Liu, Guoxuan Wang, Hongyang Yang, and Daochen Zha. Fingpt: Democratizing internet-scale data for financial large language models. *arXiv preprint arXiv:2307.10485*, 2023.
- [10] Neng Wang, Hongyang Yang, and Christina Dan Wang. Fingpt: Instruction tuning benchmark for open-source large language models in financial datasets. *arXiv preprint arXiv:2310.04793*, 2023.
- [11] George Fatouros, Kostas Metaxas, John Soldatos, and Manos Karathanassis. Marketsenseai 2.0: Enhancing stock analysis through llm agents. Papers 2502.00415, arXiv.org, Feb 2025.
- [12] George Fatouros, Kostas Metaxas, John Soldatos, and Dimosthenis Kyriazis. Can large language models beat wall street? evaluating gpt-4's impact on financial decision-making with marketsenseai. *Neural Computing and Applications*, pages 1–26, 2024.
- [13] Yi Yang, Yixuan Tang, and Kar Yan Tam. Investlm: A large language model for investment using financial domain instruction tuning. Papers 2309.13064, arXiv.org, Sep 2023.
- [14] Zhaowei Liu, Xin Guo, Fangqi Lou, Lingfeng Zeng, Jinyi Niu, Zixuan Wang, Jiajie Xu, Weige Cai, Ziwei Yang, Xueqian Zhao, Chao Li, Sheng Xu, Dezhong Chen, Yun Chen, Zuo Bai, and Liwen Zhang. Fin-r1: A large language model for financial reasoning through reinforcement learning, 2025.
- [15] Yanjun Zheng, Xiyang Du, Longfei Liao, Xiaoke Zhao, Zhaowen Zhou, Jingze Song, Bo Zhang, Jiawei Liu, Xiang Qi, Zhe Li, Zhiqiang Zhang, Wei Wang, and Peng Zhang. Agentar-fin-r1: Enhancing financial intelligence through domain expertise, training efficiency, and advanced reasoning, 2025.
- [16] Alejandro Lopez-Lira, Jihoon Kwon, Sangwoon Yoon, Jy yong Sohn, and Chanyeol Choi. Bridging language models and financial analysis, 2025.
- [17] Guojun Xiong, Zhiyang Deng, Keyi Wang, Yupeng Cao, Haohang Li, Yangyang Yu, Xueqing Peng, Mingquan Lin, Kaleb E Smith, Xiao-Yang Liu, Jimin Huang, Sophia Ananiadou, and Qianqian Xie. FLAG-TRADER: Fusion LLM-agent with gradient-based reinforcement learning for financial trading. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, *Findings of the Association for Computational Linguistics: ACL 2025*, pages 13921–13934, Vienna, Austria, July 2025. Association for Computational Linguistics.
- [18] Kemal Kirtac and Guido Germano. Sentiment trading with large language models. *Finance Research Letters*, 62:105227, 2024.
- [19] Saizhuo Wang, Hang Yuan, Lionel M. Ni, and Jian Guo. Quantagent: Seeking holy grail in trading by self-improving large language model, 2024.
- [20] Noguer I Alonso, Hanane Dupouy, et al. Evaluating llms in financial tasks-code generation in trading strategies. *Hanane, Evaluating LLMs in Financial Tasks-Code Generation in Trading Strategies (March 8, 2024)*, 2024.

- [21] Thomas Henning, Siddhartha M. Ojha, Ross Spoon, Jiatong Han, and Colin F. Camerer. Llm trading: Analysis of llm agent behavior in experimental asset markets, 2025.
- [22] Siyi Wu, Zhaoyang Guan, Leyi Zhao, Xinyuan Song, Xinyu Ying, Hanlin Zhang, Michele Pak, Yangfan He, Yi Xin, Jianhui Wang, and Tianyu Shi. Mountainlion: A multi-modal llm-based agent system for interpretable and adaptive financial trading, 2025.
- [23] Sorouralsadat Fatemi and Yuheng Hu. Finvision: A multi-agent framework for stock market prediction. In *Proceedings of the 5th ACM International Conference on AI in Finance*, ICAIF '24, page 582–590, New York, NY, USA, 2024. Association for Computing Machinery.
- [24] Yangyang Yu, Zhiyuan Yao, Haohang Li, Zhiyang Deng, Yuechen Jiang, Yupeng Cao, Zhi Chen, Jordan W. Suchow, Zhenyu Cui, Rong Liu, Zhaozhuo Xu, Denghui Zhang, Koduvayur Subbalakshmi, Guojun Xiong, Yueru He, Jimin Huang, Dong Li, and Qianqian Xie. Fincon: A synthesized llm multi-agent system with conceptual verbal reinforcement for enhanced financial decision making. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 137010–137045. Curran Associates, Inc., 2024.
- [25] Xiangyu Li, Yawen Zeng, Xiaofen Xing, Jin Xu, and Xiangmin Xu. Hedgeagents: A balanced-aware multi-agent financial trading system. In *Companion Proceedings of the ACM on Web Conference 2025*, WWW '25, page 296–305, New York, NY, USA, 2025. Association for Computing Machinery.
- [26] Tao Xu, Zhe Piao, Tadashi Mukai, Yuri Murayama, and Kiyoshi Izumi. Emulating fundamental analysts: Analytical stage-based multi-agent framework enhanced with expert guidance and preference-anchored likelihood adjustment. *Intelligent Systems with Applications*, 26:200496, 2025.
- [27] Anonymous. Automate strategy finding with LLM in quant investment. In *Submitted to ACL Rolling Review - May 2025*, 2025. under review.