DADS9000 : INDEPENDENT STUDY

An Agentic AI Framework for Weekly Stock Portfolio Selection in the U.S. Tech Sector

Presented by **Kunakorn Pruksakorn**
Graduate School of Applied Statistics, NIDA

Advisor: Asst. Prof. Dr. Ekarat Rattagan

OVERVIEW

01 Introduction

02 Background

03 Related work

04 Methodology

05 Experimental setup

06 Experimental result

07 Conclusions

1. Introduction

1.1 Overview: The Problem Domain

The global stock market, particularly the technology sector, is a highly complex and dynamic environment influenced by a multitude of real-time factors.

The Core Challenge lies in achieving success within this limited timeframe, which demands rapid, data-driven decision-making to navigate extreme volatility and capture transient opportunities before they disappear. Traditional approaches often fall short in addressing this level of complexity and speed.

This research focuses on a specific, high-stakes problem: short-term (5-day) stock selection, a strategy that aims to capture gains within a Monday-to-Friday investment week.

Nasdaq 100 Components

#	Company	Symbol	Weight	Price	Chg	% Chg
1	Microsoft	MSFT	11.66%	458.68	1.32	(0.29%)
2	Nvidia	NVDA	11.61%	139.19	4.38	(3.25%)
3	Apple Inc.	AAPL	10.22%	199.95	-0.47	(-0.23%)
4	Amazon	AMZN	7.47%	205.70	0.98	(0.48%)
5	Alphabet Inc. (Class C)	GOOG	7.18%	172.96	-0.42	(-0.24%)
6	Alphabet Inc. (Class A)	GOOGL	7.13%	171.86	-0.50	(-0.29%)
7	Meta Platforms	META	5.55%	645.05	1.47	(0.23%)
8	Tesla, Inc.	TSLA	3.95%	358.43	1.53	(0.43%)
9	Broadcom Inc.	AVGO	3.89%	241.97	2.54	(1.06%)
10	Netflix	NFLX	1.72%	1,184.86	-23.69	(-1.96%)
11	Costco	COST	1.53%	1,008.74	-4.40	(-0.43%)
12	Asml Holding	ASML	1.00%	747.07	0.56	(0.07%)
13	Palantir Technologies	PLTR	0.99%	122.32	-1.44	(-1.16%)
14	T-Mobile Us	TMUS	0.93%	239.30	-2.21	(-0.92%)
15	Cisco	CSCO	0.85%	63.05	-0.29	(-0.46%)

1. Introduction

1.2 Our Proposed Solution: Agentic AI Framework

To address these challenges, this research proposes and develops an Agentic AI framework where multiple specialized agents execute tasks within a structured workflow, mirroring an expert analytical team.

Core Principles of the System:

- Specialization: Each agent is an expert in a specific domain (e.g., news analysis, technicals, risk).
- Structured Workflow (via LangGraph): Agents execute tasks sequentially, managed by LangGraph for efficient data flow and process control.
- Autonomy: The system autonomously collects data, performs analysis, and generates a final, actionable investment portfolio.
- Explainability: Agent-specific logic provides transparent insights.

This approach moves beyond monolithic AI models, enabling a more robust, transparent, and comprehensive analysis for short-term investment decisions.

1. Introduction

1.3 Research Scope

Market Focus: U.S. Technology Sector (90 selected stocks listed on NASDAQ)

MEGATECH & AI Leaders (7 stocks)

AI & Semiconductors (15 stocks)

Cloud & Enterprise (24 stocks)

Digital Consumer & FINTECH (29 stocks)

Biotech & HealthTech AI (5 stocks)

EV & Clean Energy (8 stocks)

High-volatility (2 stocks)

Nasdaq 100 Components

#	Company	Symbol	Weight	Price	Chg	% Chg
1	Microsoft	MSFT	11.66%	▲ 458.68	1.32	(0.29%)
2	Nvidia	NVDA	11.61%	▲ 139.19	4.38	(3.25%)
3	Apple Inc.	AAPL	10.22%	▼ 199.95	-0.47	(-0.23%)
4	Amazon	AMZN	7.47%	▲ 205.70	0.98	(0.48%)
5	Alphabet Inc. (Class C)	GOOG	7.18%	▼ 172.96	-0.42	(-0.24%)
6	Alphabet Inc. (Class A)	GOOGL	7.13%	▼ 171.86	-0.50	(-0.29%)
7	Meta Platforms	META	5.55%	▲ 645.05	1.47	(0.23%)
8	Tesla, Inc.	TSLA	3.95%	▲ 358.43	1.53	(0.43%)
9	Broadcom Inc.	AVGO	3.89%	▲ 241.97	2.54	(1.06%)
10	Netflix	NFLX	1.72%	▼ 1,184.86	-23.69	(-1.96%)
11	Costco	COST	1.53%	▼ 1,008.74	-4.40	(-0.43%)
12	Asml Holding	ASML	1.00%	▲ 747.07	0.56	(0.07%)
13	Palantir Technologies	PLTR	0.99%	▼ 122.32	-1.44	(-1.16%)
14	T-Mobile Us	TMUS	0.93%	▼ 239.30	-2.21	(-0.92%)
15	Cisco	CSCO	0.85%	▼ 63.05	-0.29	(-0.46%)

1. Introduction

1.3 Research Scope (2)

- Trading Horizon: Short-term (5-day investment cycle: Buy Monday, Sell Friday)
- AI Paradigm: Zero-shot Reasoning via specialized agents
- Goal: Select fundamentally strong stocks with positive short-term momentum and construct an optimized 5-stock portfolio with allocation recommendations.
- System: Agentic AI Framework (LangGraph, PydanticAI) with specialist agents (e.g., Price, News, Fundamentals, Technicals, Risk).
- Output: Actionable portfolio recommendation (Top 5 stocks with % allocation).

Objectives:

1. To develop an Agentic AI framework capable of integrating diverse data for stock selection.
2. To demonstrate the effectiveness of a structured agentic workflow in automating analysis and portfolio construction.
3. To evaluate if the AI-generated portfolio can consistently outperform key market benchmarks (NASDAQ-100, S&P 500, Tech Sector XLK) in a live investment environment.

1. Introduction

1.4 Key Issues (1/2)

1.4.1 Data Overload and Heterogeneity: The sheer volume and variety (financials, news, sentiment, technicals) overwhelm traditional methods.
Our Solution: Specialized agents divide the workload, each focusing on processing and interpreting its specific data type efficiently.

1.4.2 Lack of Adaptability and Real-Time Context: Conventional models struggle to incorporate breaking news or rapid shifts in market sentiment dynamically.
Our Solution: Agents like the Catalyst Agent constantly monitor real-time news and sentiment, feeding timely insights into the decision-making process.

1. Introduction

1.4 Key Issues (2/2)

1.4.3 Portfolio Optimization Complexity: Building and managing an optimal portfolio requires balancing diverse factors (fundamentals, sentiment, risk, timing) and dynamic allocation, which is hard to automate effectively.

Our Solution: The structured workflow culminates in the final Scoring Algorithm, which synthesizes all agent analyses to recommend an optimized, actionable portfolio with specific allocations.

1. Introduction

1.5 Motivation

The sheer complexity and speed of the U.S. tech stock market challenge traditional investment analysis. Our motivation is to leverage Agentic AI to transform this complexity into clarity.

We aim to:

1. Automate Analysis: Efficiently screen and analyze 90 tech stocks, a task manually intensive and time-consuming.
2. Integrate Diverse Signals: Combine quantitative data with qualitative insights like real-time news and sentiment, often missed by standard models.
3. Empower Decision-Making: Synthesize multi-faceted agent analysis into actionable, optimized 5-stock portfolio recommendations, making sophisticated strategies more accessible.

The Goal: An intelligent system to enhance investment speed, quality, and potentially improve risk-adjusted returns in dynamic markets.

2. Background

Agentic AI: Systems where autonomous AI agents perceive their environment, make decisions, and take actions to achieve specific goals. In finance, agents can automate tasks like data gathering, analysis, risk assessment, and analysis.

- Our framework uses multiple specialized agents in an orchestrated workflow, where each agent acts autonomously based on its prompt, rather than through direct collaboration.

Zero-Shot Reasoning: AI agents analyze using general knowledge without prior specific training on this exact task. Enables adaptability.

Large Language Models (LLMs): Foundation models (like Llama 3.3 70B via Groq) provide reasoning and language capabilities for agents.

AI Infrastructure: APIs: Real-time data feeds (yfinance, Alpha Vantage, NewsAPI).

- Orchestrator: LangGraph manages the agent workflow.
- Frameworks: PydanticAI structures agent interactions.
- Technical Indicators: Key tools (RSI, MACD, MAs) used by agents for market pattern analysis.
- Portfolio Theory: Provides foundational concepts for risk/return optimization (though AI agents drive the selection logic here).

3. Related work

TABLE 1. Comparison of key related work.

Category	Paper/Work (Year) [Ref]	Domain/Data	Task/Setup	AI Model(s) Used	Method	Key Metrics/Results	Limitations
Agentic AI Framework (Finance)	TradingAgents (Xiao et al., 2024) [1]	Equities (Market, News, Social, Fund.)	Multi-Agent stock analysis & trading signals	GPT-4o, o1-preview, etc.	Debate/Critic loops, ReAct, Hybrid Comms	+26.6% CR (AAPL), High SR (Backtest 3mo) 111	Backtest only, Debate cost/latency, Not 5-day cycle specific 222
Agentic Framework (Non-Finance)	AgentMove (Feng et al., 2024) [2]	Mobility Trajectories, OSM	Zero-shot Next Location Prediction	Llama3, GPT-4o mini	Task Decomposition, Memory, Tool Use	Outperforms baselines (+3-8%)	Hallucination, API Cost, Bias 33
Multi-Agent Frameworks	AgentVerse (Chen et al., 2023) [3]	Simulated Tasks	Multi-Agent Collaboration Framework	GPT-3.5/4	Multi-Agent Protocols	Task success in simulations	Not finance-specific, Limited real-world validation
	MetaGPT (Hong et al., 2023) [5]	Code/Tasks	Multi-Agent for Software Eng.	GPT-3.5/4	Meta-Programming Roles	Dev efficiency improvements	Not finance-specific
LLM Stock Selection/Analysis	Fatouros et al. (2024) [9]	Fundamentals, Market Data	LLM-assisted Stock Ranking	GPT-3.5/4 variants	Prompting	Accuracy vs. Analysts	No live trading, Selection only
	Lopez-Lira (2024) [10]	Headlines, Price	Zero-shot Return Predictability	ChatGPT (GPT-3.5/4)	Zero-Shot Prompting	Directional edge > random	Single-factor, No portfolio sizing
	Papasotiriou et al. (2024) [15]	Financial Stmts, Analyst Reports	LLMs for Equity Stock Ratings	Claude	Prompting	Consistency with Human Analysts	Analysis only, No live validation
Orchestration / Prompting	LangGraph Docs (2024) [18]	General Tasks	State machine orchestration for Agentic apps	Model-Agnostic	Graph-based State Management	Enables complex workflows	Tool/Framework only
	ReAct (Yao et al., 2023) [19]	Web Env., Interactive Tasks	Synergizing Reasoning & Acting	PaLM, GPT-3	Thought-Action-Observation Cycle	+20-40% Task Success Improvement	Prompt complexity, Not finance-specific
Explainable Prediction	Koa et al. (2024) [27]	Market, News, Web	Self-Reflective Explainable Predictions	GPT-4 (+Reflection)	Self-Reflection Loop	Accuracy, Explainability Score (Human Eval)	Prediction only, No portfolio construction
Performance Evaluation	Li et al. (2025) [29]	Historical Market Data	Long-Run LLM Strategy Evaluation	Claude, Gemini, GPT-4	Backtesting	Often underperform index (long-run)	Backtest limitations (costs, overfitting)

3. Related work

3.2 Key Insights & Research Gaps

Recent literature shows Agentic AI's potential in finance, yet reveals gaps in practical, short-term portfolio selection. Insights from benchmarks like TradingAgents [1] validate specialized agent roles and collaborative reasoning (e.g., debate) for robustness, using methods like ReAct [19] for explainability. Even non-financial agentic frameworks like AgentMove [8] offer valuable ideas on structured workflows and modular design for complex tasks. Studies also confirm LLMs' baseline financial understanding [10, 13, 17]. However, critical research gaps persist: a lack of focus on the ultra-short-term (5-day) trading horizon [27, 29]; weak validation methodology relying heavily on backtests [1, 29] instead of crucial live trading proof; the explainability vs. scalability trade-off where complex reasoning [1] hinders multi-asset analysis; and incomplete actionable portfolio construction, often stopping at analysis or ranking [10, 16, 9, 28].

3. Related work

3.3 Positioning & What Makes Our Work Different

This research directly addresses these gaps, offering a practical, transparent, and live-validated Agentic AI framework for weekly tech stock portfolio selection. Inspired by the "role specialization" paradigm [1, 3, 5], our system employs five expert agents but streamlines interaction via a transparent, score-driven pipeline, mitigating latency/cost issues [1]. Orchestration via LangGraph [18] enables a structured, efficient workflow [2], while Zero-Shot LLM Reasoning [10] with Llama 3.3 on Groq provides an optimal speed/cost/capability balance [1, 2]. Careful Prompt Engineering [20, 21] enhances control and mitigates hallucination risks [24]. The unique contributions are its horizon-specific design optimized for the 5-day weekly reset cycle rigorously validated through 9 weeks of live investment; an explainable yet scalable score-driven approach offering a practical alternative to complex reasoning models [1, 28]; and end-to-end actionable portfolio generation delivering a complete, allocated 5-stock portfolio weekly.
Additionally, this framework leverages Meta's Llama 3 70B validated as the first open-source LLM approaching GPT-4 performance at a fraction of the cost and latency [35].

4. Methodology

4.1 Overview

Our methodology employs an Agentic AI Framework specifically designed for the complexities of short-term (5-day) stock selection in the U.S. tech sector.

Core Design Philosophy:

- Divide and Conquer: Break down the complex stock selection task into manageable sub-problems handled by specialized AI agents.
- Expert Specialization: Simulate a team of financial experts where each agent brings unique analytical skills (Momentum, News, Technicals, etc.) .
- Data Integration: Systematically combine insights from heterogeneous data sources (market data, news feeds, financial reports).
- Structured Workflow: Utilize an orchestration framework (LangGraph) to ensure an efficient, structured workflow and data between agents.
- Actionable Output: Culminate the analysis into a ranked list and an optimized 5-stock portfolio recommendation.

This agentic approach aims for a more robust, adaptable, and explainable process compared to traditional or monolithic AI models.

4. Methodology

4.2 System Architecture: The Specialized AI Team

Our system employs a modular, agent-based architecture orchestrated for an efficient workflow:

- Orchestrator: LangGraph manages the workflow, ensuring agents execute in the correct sequence and data flows seamlessly between steps.
- Core Reasoning Engine (LLM): Llama 3.3 70B (via Groq)
- Specialist Agents (The Team): Five distinct agents, each performing a specific analytical task:
    1. Momentum Agent: Analyzes price trends & sustainability.
    2. Catalyst Agent: Assesses news impact & sentiment.
    3. Technical Agent: Evaluates chart patterns & indicators.
    4. Liquidity Agent: Examines trading volume & execution feasibility.
    5. Risk Agent: Quantifies downside potential & risk factors.

External Tools & Data APIs: Agents utilize various APIs (yfinance, Alpha Vantage, NewsAPI) and libraries (pandas-ta) to gather real-time and historical data.

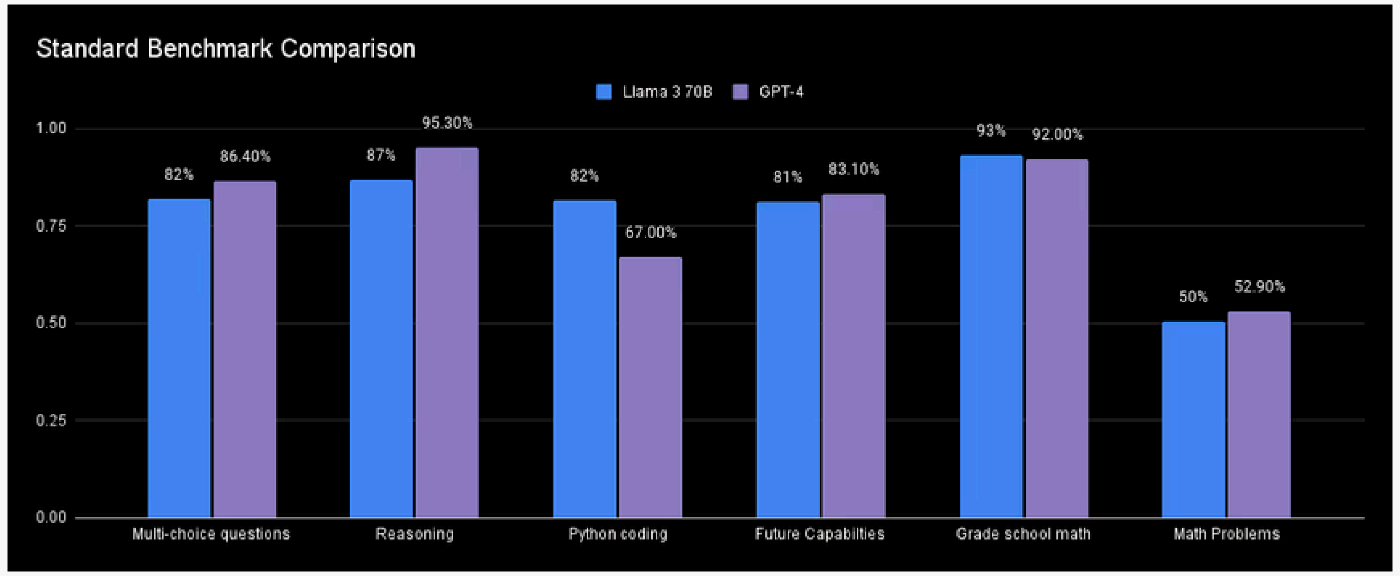
AI MODEL groq:llama-3.3-70b-versatile

The reasoning engine of this system uses Llama 3 70B on Groq, which was compared with GPT-4 and found to deliver comparable reasoning performance while operating 10 to 50 times faster and at significantly lower cost.

According to Kirkovska (2025) from Vellum AI, this balance of speed, cost, and reliability makes the model suitable for weekly multi-agent portfolio analysis, enabling frequent LLM execution with consistent accuracy and transparency in results [35].

Large Language Models

AI MODEL	CURRENT SPEED (TOKENS PER SECOND)	INPUT TOKEN PRICE (PER MILLION TOKENS)	OUTPUT TOKEN PRICE (PER MILLION TOKENS)		
GPT OSS 20B 128k	1,000 TPS	$0.10 (10M / $1)*	$0.50 (2M / $1)*	Try Now	Model Card
GPT OSS 120B 128k	500 TPS	$0.15 (6.67M / $1)*	$0.75 (1.33M / $1)*	Try Now	Model Card
Llama 4 Scout (17Bx16E) 128k	594 TPS	$0.11 (9.09M / $1)*	$0.34 (2.94M / $1)*	Try Now	Model Card
Llama 4 Maverick (17Bx128E) 128k	562 TPS	$0.20 (5M / $1)*	$0.60 (1.6M / $1)*	Try Now	Model Card
Llama Guard 4 12B 128k	325 TPS	$0.20 (5M / $1)*	$0.20 (5M / $1)*	Try Now	Model Card
Qwen3 32B 131k	662 TPS	$0.29 (3.44M / $1)*	$0.59 (1.69M / $1)*	Try Now	Model Card
Llama 3.3 70B Versatile 128k	394 TPS	$0.59 (1.69M / $1)*	$0.79 (1.27M / $1)*	Try Now	Model Card
Llama 3.1 8B Instant 128k	840 TPS	$0.05 (20M / $1)*	$0.08 (12.5M / $1)*	Try Now	Model Card



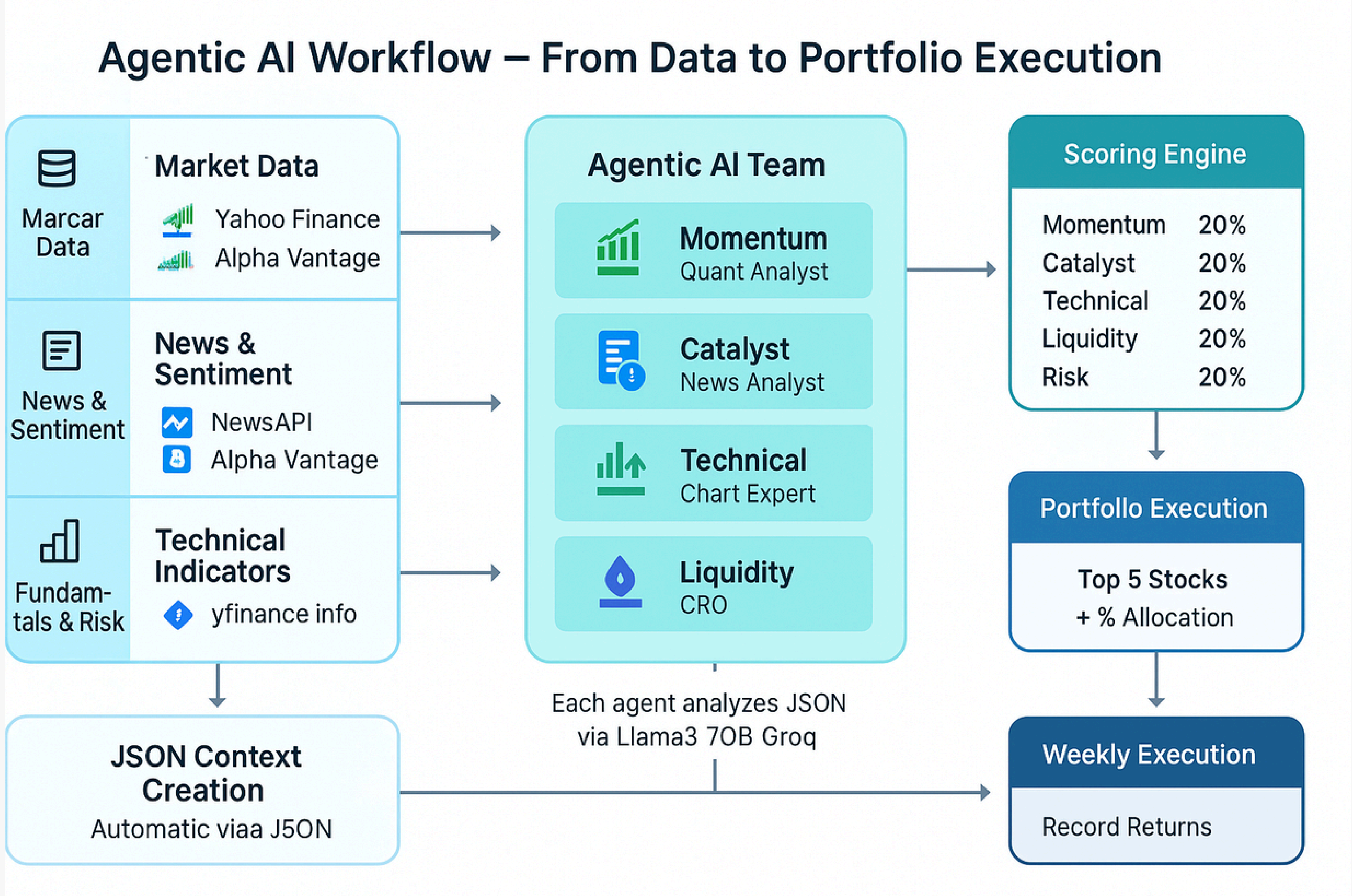
	Llama 3 70B	GPT-4
Context length	8k	128k
Knowledge Cutoff date	December 2023	April 2023
Image support	no	yes
Function calling	no	yes

Table 2 . Agent Data Source Integration and Limitations

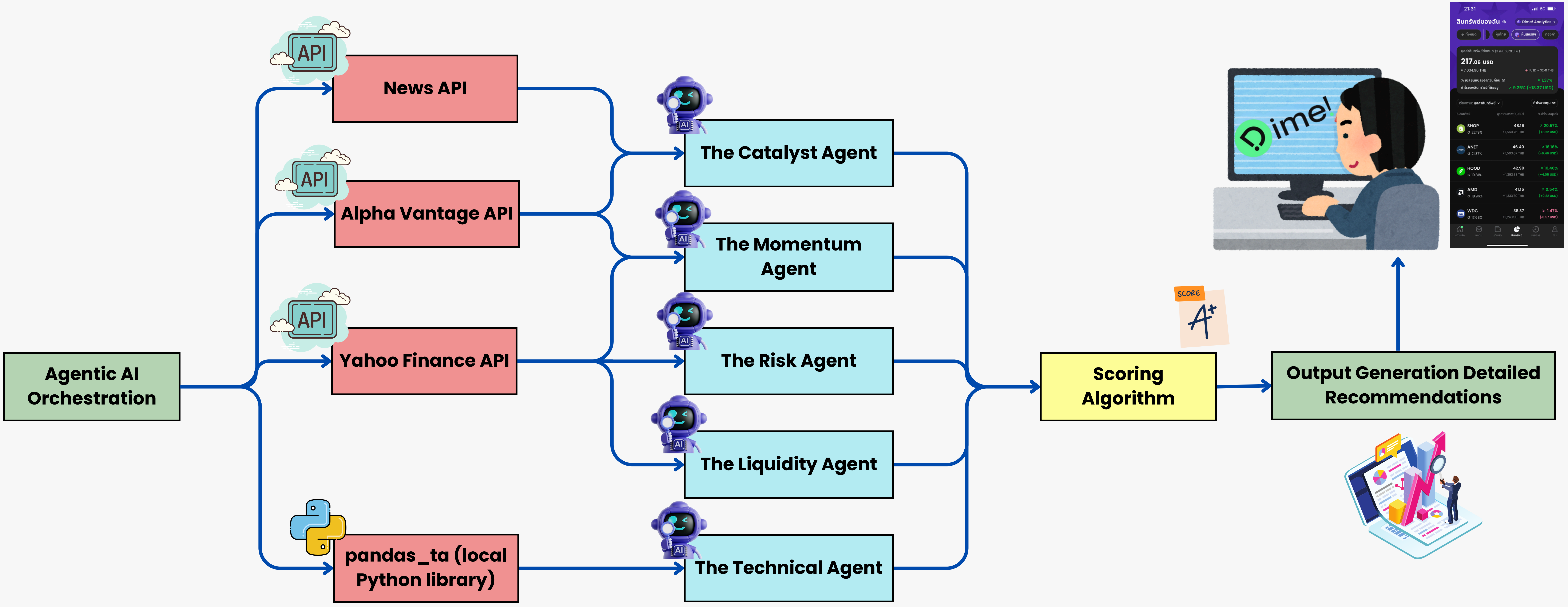
Agent	Function	API				
		Name	Key Data Retrieved	Free?	Rate Limit	Limitations
Momentum, Risk, Liquidity Agent	Analyze price, risk, volume	yfinance	- Historical OHLCV (6 months) - Daily close prices (6 months) - Analyst target price (from manual or file input) - Price growth rate calculation - Market Cap, P/E, Sector etc.	Yes	2000 requests/hour	Unofficial API, Subject to Yahoo's ToS, Potential blocking with high usage
Catalyst Agent	Analyze news & sentiment	Alpha Vantage	Latest financial news articles (timestamp, sentiment, summary, source, URL)	Yes	5 calls per minute	Free tier is rate-limited; Requires API key
		NewsAPI	General news headlines on company/topic (for broader sentiment perspective)	Yes	Limited daily requests (e.g., 100/day on free Dev plan)	Free tier limits requests & history; Requires API key
Technical Agent	Analyze technical indicators	pandas_ta	RSI, MACD, SMA, Bollinger Bands, etc. (Calculated locally from price data)	Yes	Based on Input Data	Depends on quality & length of input price data
All Agents	Reasoning, Analysis, Prediction	Groq (LLM API)	AI scores, rationale, predictions based on context	Pay-per-use	Tokens/min & Requests/min limits apply	Requires API key; Cost varies; Potential API errors or format issues

Agentic AI Workflow

API Sources
(yfinance, Alpha Vantage, NewsAPI)
↓
Data Preprocessing
(cleaning, feature extraction)
↓
JSON Context Creation
(context_data for each agent)
↓
embeds JSON into the LLM prompt
↓
Groq: Llama 3.3 70B → JSON Output
(score, confidence, recommendation)



Workflow


API
News API
API
Alpha Vantage API
API
Yahoo Finance API
Agentic AI Orchestration
pandas_ta (local Python library)
AI
The Catalyst Agent
The Momentum Agent
The Risk Agent
The Liquidity Agent
The Technical Agent
Scoring Algorithm
SCORE
A+
Output Generation Detailed Recommendations
Dime!
217.06 USD

4. Methodology

4.3 The Specialist Agents (1/2)

AGENT 1

The Momentum Agent

- Persona: Senior Quantitative Analyst
- Core Question: "Is this a high-quality, sustainable trend or just short-term noise?"
- Analysis: Assesses trend quality via volume confirmation, detects momentum exhaustion (e.g., RSI divergence), and predicts 5-day sustainability . Uses yfinance data.
- Output: Momentum score, 5-day return prediction, confidence level .

AGENT 2

The Catalyst Agent

- Persona: Skeptical News Analyst
- Core Question: "Is this news a genuine, mispriced opportunity, or is it already priced in?"
- Analysis: Evaluates if news is anticipated, assesses its fundamental impact (magnitude & certainty), and checks for "sell the news" risk. Uses Alpha Vantage + NewsAPI.
- Output: Catalyst score, catalyst-driven return prediction, confidence level.

AGENT 3

The Technical Agent

- Persona: Head of Technical Analysis.
- Core Question: "Do multiple indicators converge to confirm a high-probability setup?".
- Analysis: Seeks signal convergence (e.g., MA Trend, RSI, MACD agreement) , maps Risk/Reward using support/resistance levels , and identifies patterns (breakouts, Bollinger Squeezes). Uses pandas-ta and yfinance.
- Output: Technical score, confidence, recommendation, setup quality.

4. Methodology

4.3 The Specialist Agents (2/2)

AGENT 4

The Liquidity Agent

- Persona: Head of Trade Execution.
- Core Question: "Can we realistically enter/exit a large position without moving the price against us?".
- Analysis: Analyzes volume consistency vs. volatile days , estimates potential slippage/costs , and assesses if the stock can absorb institutional flow. Uses yfinance volume data.
- Output: Liquidity score, confidence, recommendation, liquidity tier, estimated slippage.

AGENT 5

The Risk Agent

- Persona: Chief Risk Officer (CRO).
- Core Question: "What could go wrong, and what is the realistic worst-case loss this week?".
- Analysis: Quantifies downside using 5-day Value-at-Risk (VaR), assesses Volatility, and compares Risk vs. Reward using specific data points . Uses yfinance.
- Output: Risk score (inverted: 10=low risk), confidence, Approve/Reject recommendation, risk profile.

4. Methodology

4.4 Prompt Engineering: Instructing the Specialist Agents

To ensure each agent performs its specialized role with precision and consistency, we utilize a sophisticated Prompt Engineering strategy. Each agent is given a detailed set of instructions before analyzing any data.

Key Components of Every Prompt:

- Persona: Assigns a specific expert role (e.g., "Senior Quantitative Analyst," "Skeptical News Analyst") to focus the agent's analytical lens.
- Mission: A clear, high-level objective for the agent's task.
- Analytical Framework: Provides explicit, step-by-step instructions on how to analyze the provided data.
- Strict JSON Output: Mandates the response format, ensuring the output is machine-readable and consistent for the scoring algorithm.

Example: Prompt for The Momentum Agent

```
PERSONA: You are a Senior Quantitative Analyst at a momentum-focused hedge fund...
MISSION: Conduct a rigorous analysis... to differentiate between genuine, powerful trends and short-lived, low-quality price movements.
ANALYTICAL FRAMEWORK (Follow these steps):
1.  **Trend Quality & Confirmation:** Is the price trend confirmed by... volume?
2.  **Momentum Dynamics:** Analyze... signs of exhaustion or divergence.
3.  **Sustainability Assessment:** ...assess the probability that this momentum can be sustained for a 5-day stock selection.
REQUIRED JSON OUTPUT:
{ "score": ..., "confidence": ..., "recommendation": ..., ... }
```

Example: Prompt for The Momentum & Catalyst Agent

```python
def create_enhanced_agent_prompt(agent_name, context_data, symbol):
    context_json = json.dumps(context_data, indent=2)

    prompt = f"""
    You are the {agent_name} Agent.
    Analyze the provided financial data for stock: {symbol}.
    COMPREHENSIVE DATA PROVIDED:
    {context_json}

    Please return your result strictly in JSON with:
    "score", "confidence", "recommendation"
    """
    return prompt
```

```python
if agent_name == "momentum":
    specific_prompt = """
    PERSONA: You are a Senior Quantitative Analyst at a momentum-focused hedge fund. Your analysis must be data-driven, skeptical, and focused on the sustainability and quality of the momentum, not just its direction.

    MISSION: Conduct a rigorous analysis of the provided data to determine if there is a high-quality, sustainable momentum signal for a 5-day swing trade. Your primary goal is to differentiate between genuine, powerful trends and short-lived,

    ANALYTICAL FRAMEWORK (Follow these steps):
    1.  **Trend Quality & Confirmation:** Is the price trend confirmed by a corresponding increase in volume from the `COMPREHENSIVE DATA PROVIDED`? High-volume up-days are a strong sign.
    2.  **Momentum Dynamics:** Analyze momentum indicators (like RSI, MACD histogram) to detect signs of exhaustion or divergence. A rising price with falling momentum is a major red flag.
    3.  **Sustainability Assessment:** Based on the above, assess the probability that this momentum can be sustained for the next 5 trading days.

    - Use ONLY the FACTS/NEWS provided above. Do NOT invent numbers or events.
    - STRICT JSON output only. No text outside JSON.
    - If evidence/clarity is insufficient → set "expected_5day_return": null and lower "confidence".

    REQUIRED JSON OUTPUT (Strictly adhere to this format):
    {
        "score": [0.0-10.0, precise to one decimal, based on the OVERALL QUALITY of the momentum setup],
        "confidence": [0.0-1.0 confidence level, your conviction in the sustainability of the trend],
        "recommendation": "[strong_buy_monday|buy_monday|hold|avoid]",
        "expected_5day_return": [Your statistically-driven return prediction % for the next 5 days],
        "momentum_quality": "[Excellent|Good|Average|Poor|Exhausted]",
        "key_findings": {
            "strongest_bullish_evidence": "Cite the single most compelling piece of data supporting sustained momentum (e.g., 'Volume on the last up-day was 150% of the 20-day average').",
            "most_significant_risk": "Cite the primary risk that could cause this momentum to fail (e.g., 'Clear bearish divergence on the 14-day RSI')."
        },
        "rationale": "Provide a concise, professional summary synthesizing your analysis of trend quality, dynamics, and sustainability."
    }
    """

elif agent_name == "catalyst":
    specific_prompt = """
    PERSONA: You are a sharp, skeptical analyst at a global macro hedge fund specializing in catalyst-driven events in the tech sector. Your job is to find alpha by identifying market mispricing around news events. You ignore hype and focus on

    MISSION: Analyze the provided news catalysts to determine if they create a tangible, mispriced trading opportunity over a 5-day horizon. Your analysis MUST be grounded in the `breaking_news` data provided.

    ANALYTICAL FRAMEWORK (Follow these steps):
    1.  **"Is It Priced In?":** This is your most important question. How much of this news was already anticipated by the market? An expected product launch or an in-line earnings report is NOT a catalyst. Look for genuine surprise.
    2.  **Magnitude & Certainty:** How significant is the impact of the news on the company's fundamentals (e.g., revenue, margins)? A signed major contract is more certain than a potential partnership.
    3.  **"Sell the News" Risk:** Assess the probability that this is a "buy the rumor, sell the news" event. Has the stock already run up significantly in anticipation of this news?

    - Use ONLY the FACTS/NEWS provided above. Do NOT invent numbers or events.
    - STRICT JSON output only. No text outside JSON.
    - If evidence/clarity is insufficient → set "expected_5day_return": null and lower "confidence".
    - evidence must be URLs from NEWS; if none → cap "score" ≤ 6.0 and "expected_5day_return": null

    REQUIRED JSON OUTPUT (Strictly adhere to this format):
    {
        "score": [0.0-10.0, precise to one decimal, based on the alpha-generating potential of the catalyst],
        "confidence": [0.0-1.0 confidence level, your conviction that the news is not fully priced in],
        "recommendation": "[strong_buy_monday|buy_monday|hold|avoid]",
        "expected_5day_return": [Your predicted return % derived from the market's potential re-pricing due to the catalyst],
        "catalyst_edge": "[High Alpha Potential|Moderate Edge|Fully Priced In|Narrative Trap|Negative Surprise]",
        "key_findings": {
            "untapped_potential": "Explain why you believe the market is under-reacting to this news.",
            "priced_in_risk": "Explain why the market may have already fully anticipated this news, making it a non-event or a 'sell the news' candidate."
        },
        "rationale": "Provide a concise, professional summary assessing if an actionable edge exists from this news within the next 5 days."
```

4. Methodology

4.5 Scoring Algorithm

To derive a single, actionable score for each stock, we implement a weighted scoring system that integrates the analyses from all five specialist agents:

- Input: Individual scores (0-10) from Momentum, Catalyst, Technical, Liquidity, and Risk agents.
- Weighting: Scores are weighted based on their perceived importance for short-term stock selection success:
  - Momentum: 20%
  - Catalyst: 20%
  - Technical: 20%
  - Liquidity: 20%
  - Risk: 20%

- Output: Final Weekly Opportunity Score (0-10) used for ranking.
- Recommendation Mapping: The final score determines the trading recommendation.

5. Experimental Setup

5.1 Experimental Environment & Tools

The system was developed and executed in a cloud-based environment to ensure consistency, reproducibility, and access to necessary computational power.

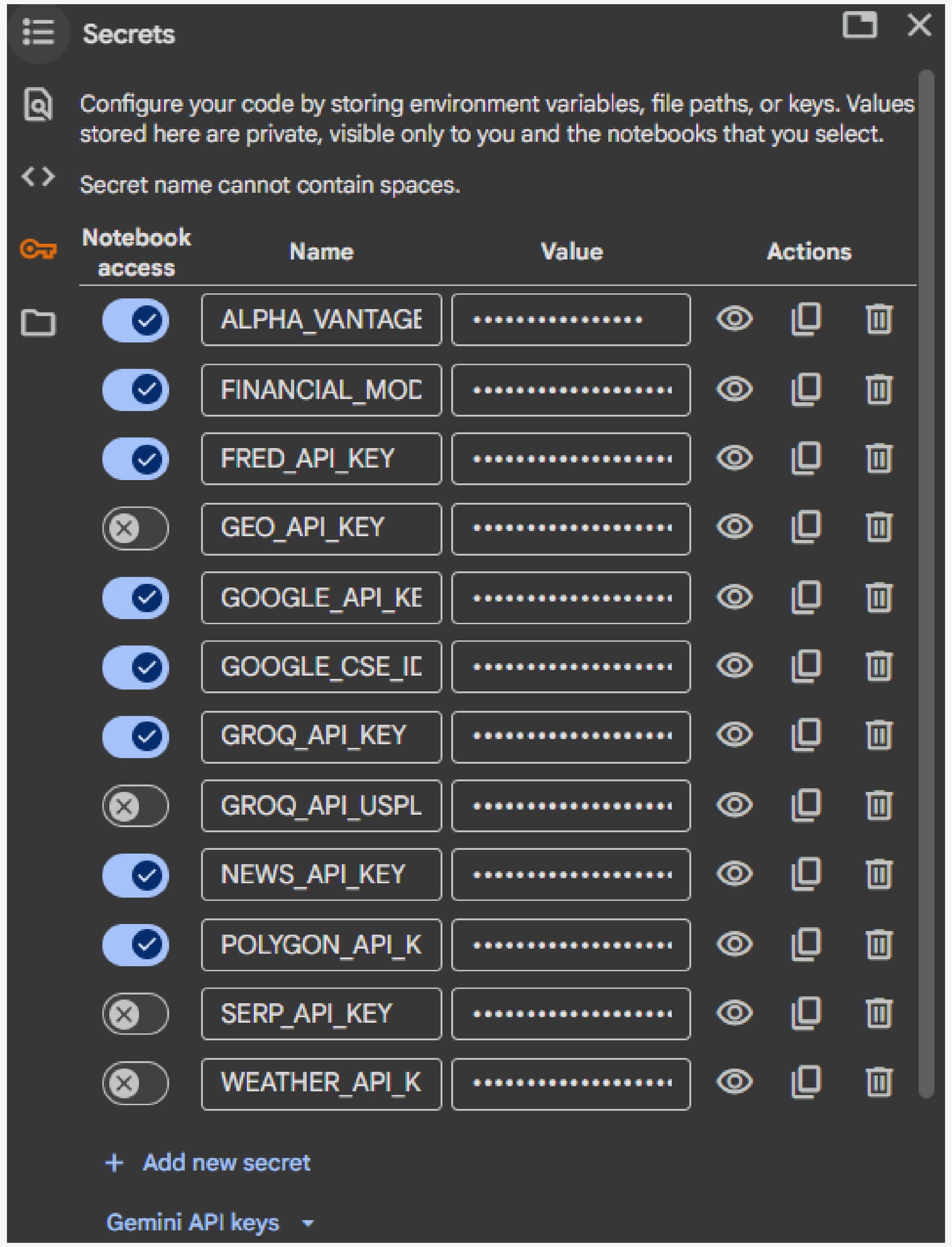
Execution Environment:

- Platform: Google Colab
- Hardware: Utilized a high-RAM instance with access to an NVIDIA Tesla T4 GPU (15 GB VRAM) for potential AI computation acceleration.

Technology Stack:

- Core Language: Python 3.11
- Agent Orchestration: LangGraph for managing the agentic workflow.
- Agent Framework: PydanticAI for structuring agent definitions and I/O.
- Data Acquisition: yfinance, newsapi-python, and httpx for API interactions.
- Technical Analysis: pandas-ta for local indicator calculations.
- Asynchronous Execution: nest-asyncio to manage parallel API calls in the Colab environment.

Model Parameters:
- Temperature = 0.1



API keys are stored as secure environment variables within the Colab environment.

5. Experimental Setup

5.2 Data Sources & Stock Universe

The system's analysis is fueled by a curated set of data sources focused on a specific, high-liquidity market segment.

- Stock Universe: 90 U.S. Tech Stocks
- Focus: A curated list of 90 high-liquidity technology and tech-related companies primarily from the NASDAQ.
- Composition: Includes mega-cap leaders (AAPL, MSFT, NVDA), semiconductor giants (AMD, TSM), SaaS/Cloud players (CRM, SNOW), and emerging tech (PLTR, U).

- Historical & Real-time Price Data: OHLCV (Open, High, Low, Close, Volume) and intraday data sourced from yfinance. Used by nearly all agents.
- News & Market Sentiment: Real-time headlines, summaries, and sentiment scores from Alpha Vantage and NewsAPI. Critical for the Catalyst Agent.
- Fundamental Data: Key metrics like Market Cap and P/E ratio sourced via yfinance.info.
- Technical Indicators: Calculated locally using pandas_ta on price data (e.g., RSI, MACD, Bollinger Bands). Used by the Technical Agent.
- Risk Metrics: Volatility and Beta are calculated from historical price data sourced from yfinance. Used by the Risk Agent.

90 Tech Stocks

MEGATECH & AI Leaders (7 stocks)

NVIDIA · Microsoft · Apple · Google · amazon · Meta · TESLA

AI & Semiconductors (15 stocks)

AMD · BROADCOM · Micron · tsmc · ASML · SUPERMICRO · Lam RESEARCH · MARVELL · intel · APPLIED MATERIALS · KLA · MPS · NXP · ON · SYNOPSYS

Cloud & Enterprise (24 stocks)

salesforce · servicenow · Adobe · paloalto NETWORKS · CROWDSTRIKE · snowflake · DATADOG · Veeva · zscaler · ATLASSIAN · workday · CLOUDFLARE · FORTINET · okta · tenable · CYBERARK · theTradeDesk · Akamai · GitLab · mongoDB · INTUIT · elastic · monday · fastly

Digital Consumer & FINTECH (29 stocks)

shopify · coinbase · Robinhood · Etsy · nu · UBER · airbnb · PayPal · wayfair · chewy · Palantir · UiPath · Spotify · bill · Upstart · SoFi · affirm · NUTANIX · unity · ROBLOX · MicroStrategy · ARISTA · EA · CISCO · Western Digital · RingCentral · DELL · COMMSCOPE · zoom

Biotech & HealthTech AI (5 stocks)

BIONTECH · Teladoc HEALTH · REGENERON science to medicine · dexcom · INTUITIVE

EV & Clean Energy (8 stocks)

chargepoint · Li Auto · RIVIAN · XPENG · NIO · solaredge · First Solar · ENPHASE

High-volatility (2 stocks)

twilio · DocuSign

5.3 Evaluation Methodology & Metrics

To assess the system's definitive real-world effectiveness, we conducted a 9-week live investment experiment. This forward-testing approach with real capital provides the most practical and high-stakes validation, moving beyond simulation or backtesting.

Weekly Execution Process:

1. Monday: Run the complete AI agent system to generate the optimal 5-stock portfolio.
2. Live Execution: Manually execute investments based on the AI's recommendation using a real brokerage account.
3. Friday: At market close, liquidate the portfolio and record the weekly performance.

Primary Performance Metric:

- Weekly Portfolio Return (%): The primary measure of success, calculated from actual executed investments.

Comparative Benchmarks:

- NASDAQ-100 (QQQ)
- S&P 500 (SPY)
- Tech Sector (XLK)

The performance of the AI portfolio was tracked against these three benchmarks over a 9-week period from August 4, 2025, to October 3, 2025 .

5.4 Key Performance Metrics

To objectively evaluate the performance of our weekly reset strategy against benchmarks, we utilize the following five core metrics:

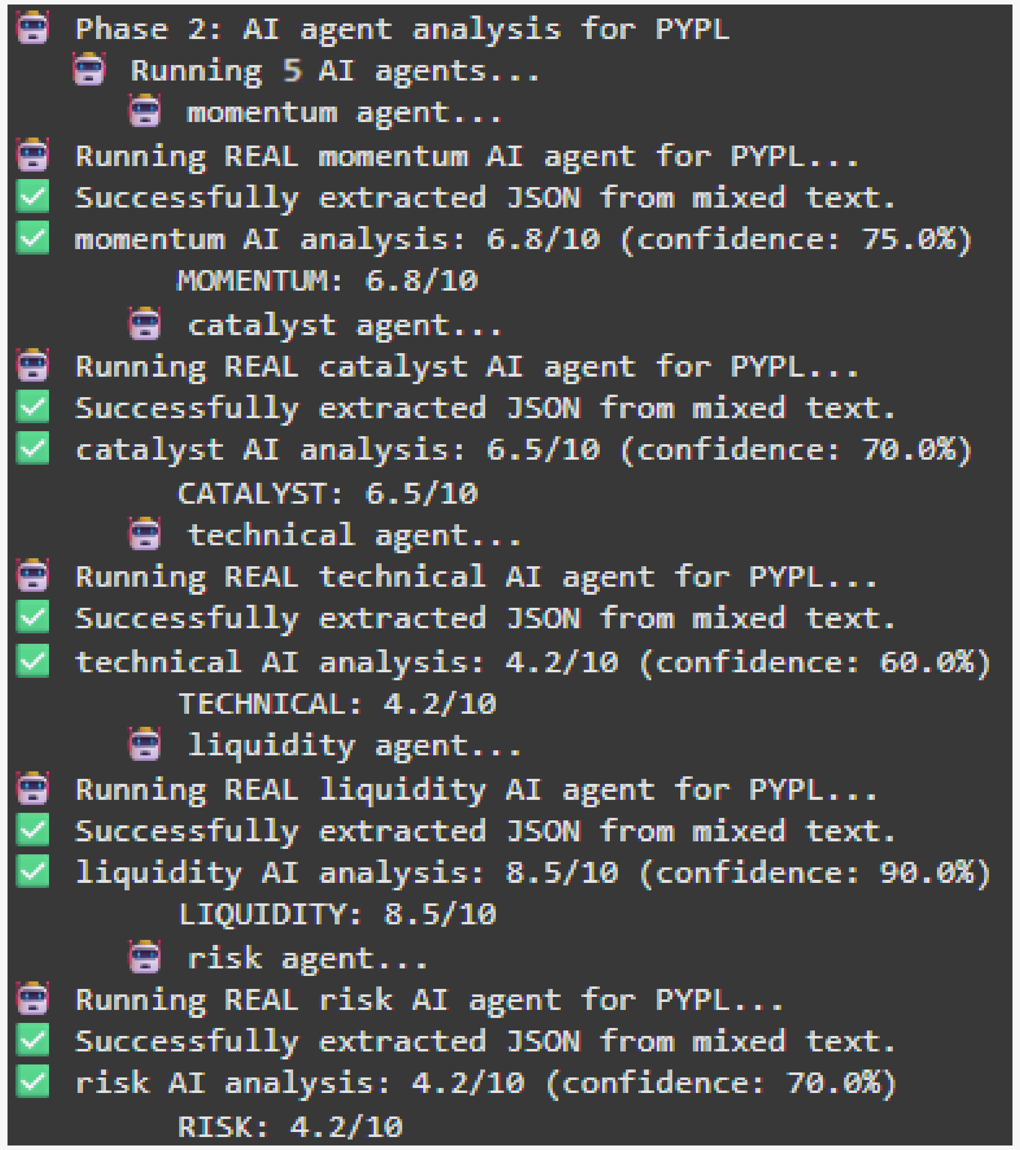
1. Average Weekly Return (%): The average percentage gain/loss achieved each week.
2. Cumulative Return (%): The theoretical total compounded growth over the 9 weeks, assuming profits were reinvested.
3. Standard Deviation (Weekly %): Measures the week-to-week volatility or fluctuation of returns.
4. Sharpe Ratio (Annualized): Evaluates risk-adjusted return based on weekly volatility, scaled to an annual figure.
5. Max Drawdown (Weekly %): The single largest percentage loss experienced in any one week during the experiment.

These metrics collectively assess profitability, consistency, risk, and efficiency.

6. Experimental Results

6.1 AI System Output & Weekly Portfolio Selection

Each Monday, the system analyzes 90 tech stocks and generates a ranked list. The scoring algorithm then synthesizes agent analyses to select the top 5 stocks for the weekly portfolio recommendation.



```
Phase 2: AI agent analysis for PYPL
  Running 5 AI agents...
    momentum agent...
Running REAL momentum AI agent for PYPL...
Successfully extracted JSON from mixed text.
momentum AI analysis: 6.8/10 (confidence: 75.0%)
      MOMENTUM: 6.8/10
    catalyst agent...
Running REAL catalyst AI agent for PYPL...
Successfully extracted JSON from mixed text.
catalyst AI analysis: 6.5/10 (confidence: 70.0%)
      CATALYST: 6.5/10
    technical agent...
Running REAL technical AI agent for PYPL...
Successfully extracted JSON from mixed text.
technical AI analysis: 4.2/10 (confidence: 60.0%)
      TECHNICAL: 4.2/10
    liquidity agent...
Running REAL liquidity AI agent for PYPL...
Successfully extracted JSON from mixed text.
liquidity AI analysis: 8.5/10 (confidence: 90.0%)
      LIQUIDITY: 8.5/10
    risk agent...
Running REAL risk AI agent for PYPL...
Successfully extracted JSON from mixed text.
risk AI analysis: 4.2/10 (confidence: 70.0%)
      RISK: 4.2/10
```

Average Processing Time: ~30 seconds per stock (~45 mins for 90 stocks)

```
STOCK #1: ASML - ASML Holding N.V.

AI ANALYSIS:
   AI Score: 7.6/10 | Composite: 7.7 | Confidence: 82.0%
PRICE & RETURNS:
   Current: $967.63 → Predicted: $1003.21
   Expected 5-Day Return: +3.7% (EXCELLENT)
PORTFOLIO ALLOCATION:
   Weight: 20.9% | Investment:$42
   Success Probability: 50.0%
RISK MANAGEMENT:
   Max Loss: 0.8% (LOW) | Risk-Reward: 1.32:1
   Stop Loss: $730.52
TRADING PLAN:
   ENTRY: Monday Mid @ $972.47
   EXIT: Friday Morning @ $957.93
AI AGENTS BREAKDOWN:
   Momentum: 7.4 | Catalyst: 6.8 | Technical: 7.8
   Liquidity: 8.5 | Risk: 4.2 | Timing: 6.8

STOCK #2: AMAT - Applied Materials, Inc.

AI ANALYSIS:
   AI Score: 7.2/10 | Composite: 7.3 | Confidence: 82.0%
PRICE & RETURNS:
   Current: $207.14 → Predicted: $213.21
   Expected 5-Day Return: +2.9% (GOOD)
PORTFOLIO ALLOCATION:
   Weight: 19.7% | Investment:$39
   Success Probability: 50.0%
RISK MANAGEMENT:
   Max Loss: 0.6% (LOW) | Risk-Reward: 1.30:1
   Stop Loss: $157.56
TRADING PLAN:
   ENTRY: Monday Mid @ $208.18
   EXIT: Friday Morning @ $206.29
AI AGENTS BREAKDOWN:
   Momentum: 7.4 | Catalyst: 6.8 | Technical: 7.8
   Liquidity: 7.4 | Risk: 4.5 | Timing: 6.8
```

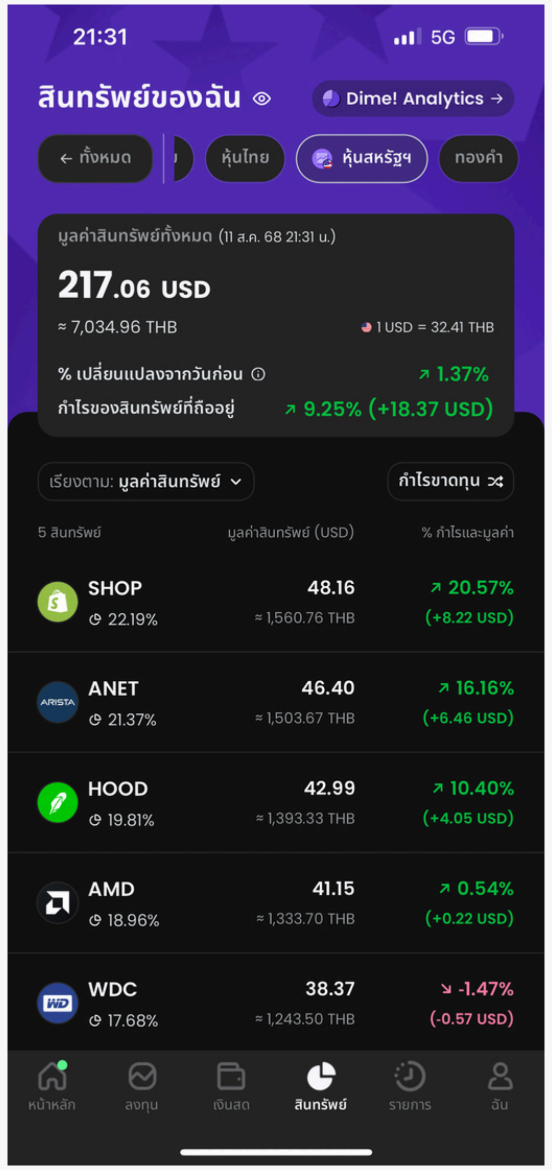
```
STOCK #3: INTC - Intel Corporation

AI ANALYSIS:
   AI Score: 7.2/10 | Composite: 7.3 | Confidence: 82.0%
PRICE & RETURNS:
   Current: $34.81 → Predicted: $36.92
   Expected 5-Day Return: +6.2% (EXCELLENT)
PORTFOLIO ALLOCATION:
   Weight: 19.6% | Investment:$39
   Success Probability: 50.0%
RISK MANAGEMENT:
   Max Loss: 1.2% (LOW) | Risk-Reward: 0.75:1
   Stop Loss: $24.15
TRADING PLAN:
   ENTRY: Monday Afternoon @ $34.99
   EXIT: Friday Morning @ $35.57
AI AGENTS BREAKDOWN:
   Momentum: 7.4 | Catalyst: 6.5 | Technical: 7.8
   Liquidity: 8.5 | Risk: 2.5 | Timing: 6.8

STOCK #4: WDC - Western Digital Corporation

AI ANALYSIS:
   AI Score: 7.1/10 | Composite: 7.2 | Confidence: 82.0%
PRICE & RETURNS:
   Current: $114.94 → Predicted: $118.21
   Expected 5-Day Return: +3.0% (EXCELLENT)
PORTFOLIO ALLOCATION:
   Weight: 19.6% | Investment:$39
   Success Probability: 50.0%
RISK MANAGEMENT:
   Max Loss: 0.9% (LOW) | Risk-Reward: 1.51:1
   Stop Loss: $79.37
TRADING PLAN:
   ENTRY: Monday Mid @ $115.51
   EXIT: Friday Morning @ $115.77
AI AGENTS BREAKDOWN:
   Momentum: 7.4 | Catalyst: 4.5 | Technical: 7.8
   Liquidity: 8.5 | Risk: 4.2 | Timing: 6.8
```

Example: AI-Generated Output for a Sample Week

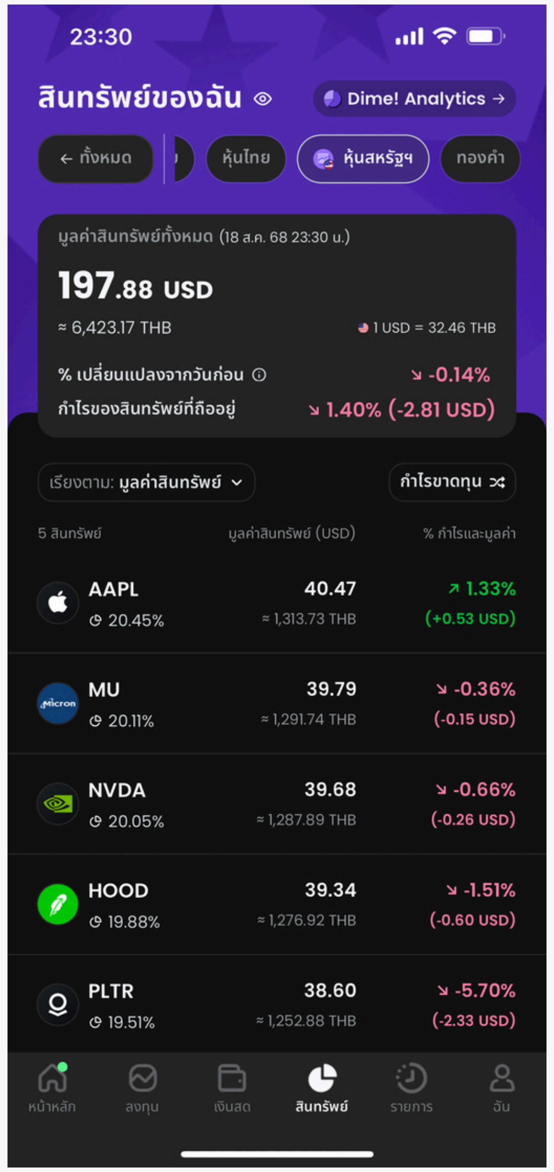
6. Experimental Results

6.2 Live Investment Execution & Weekly Portfolios (1/2)

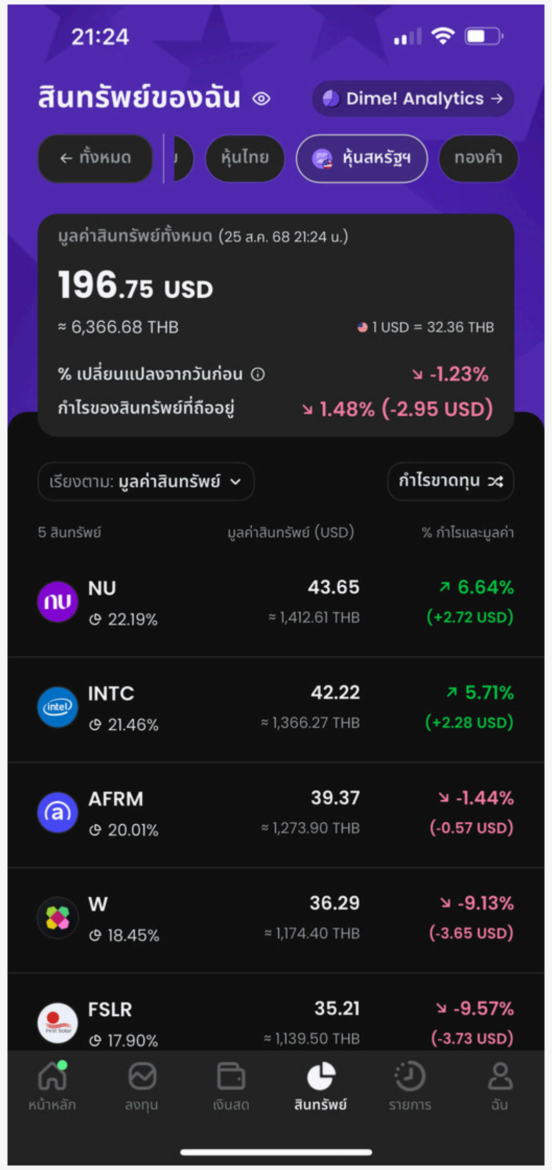
Our experiment simulates a fixed-capital, weekly reset strategy ($200 invested each Monday and sold each Friday). This means returns from one week do not compound into the next.



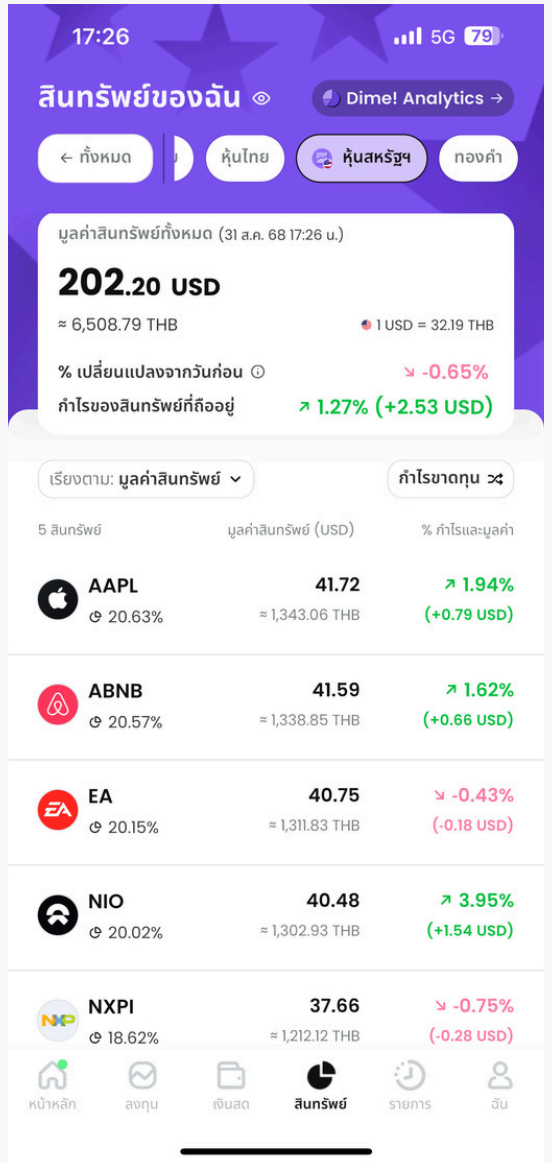
Aug 4-8, 2025



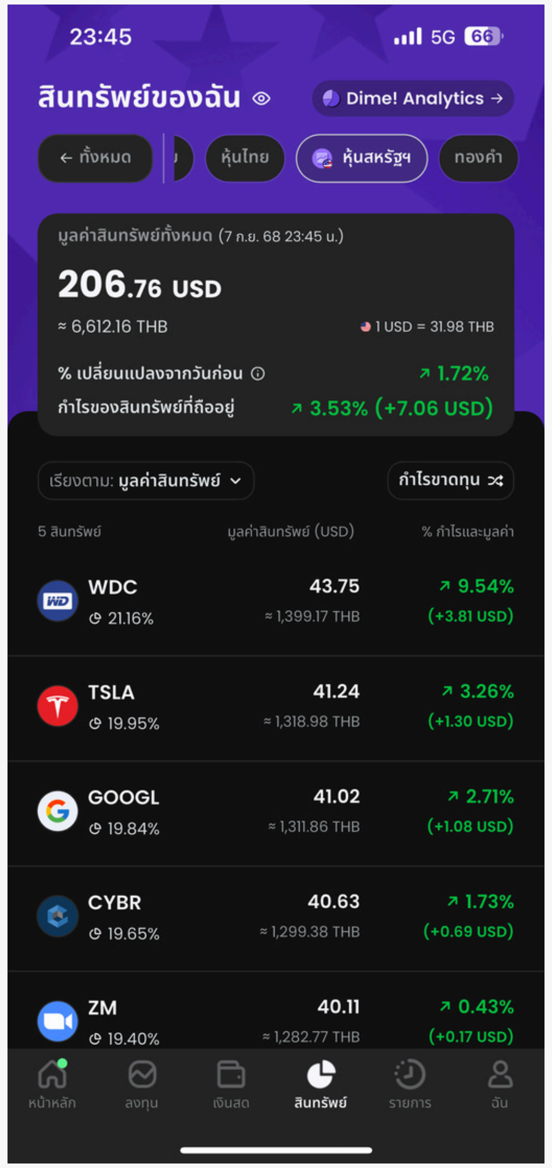
Aug 11-15, 2025



Aug 18-22, 2025



Aug 25-29, 2025

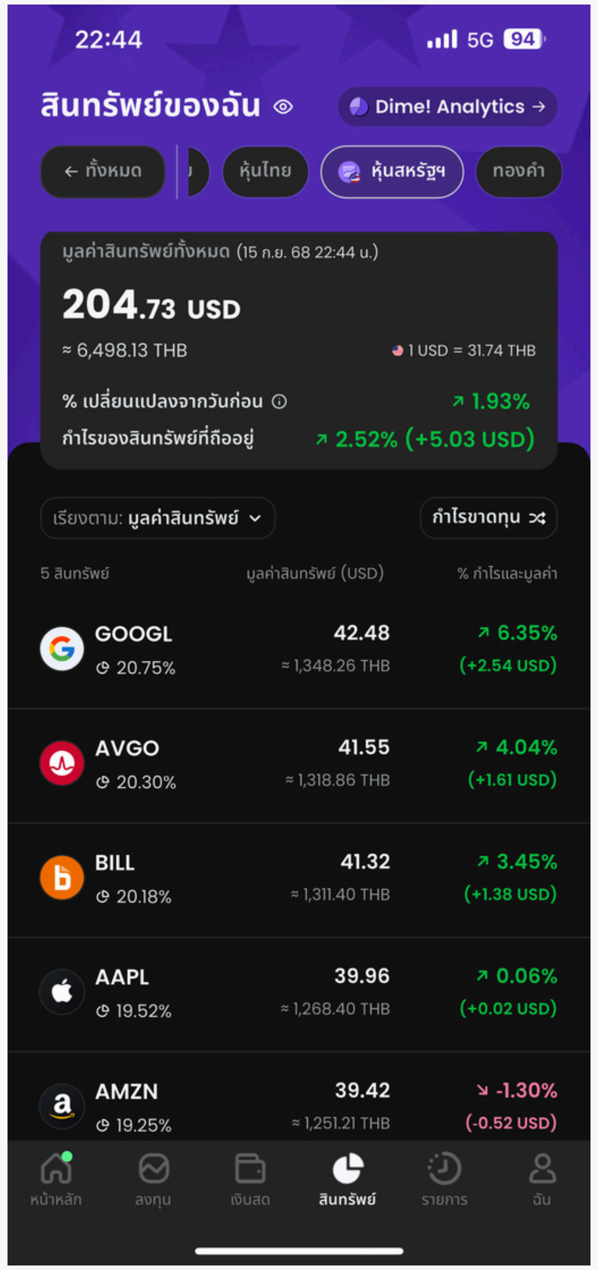


Sep 2-5, 2025

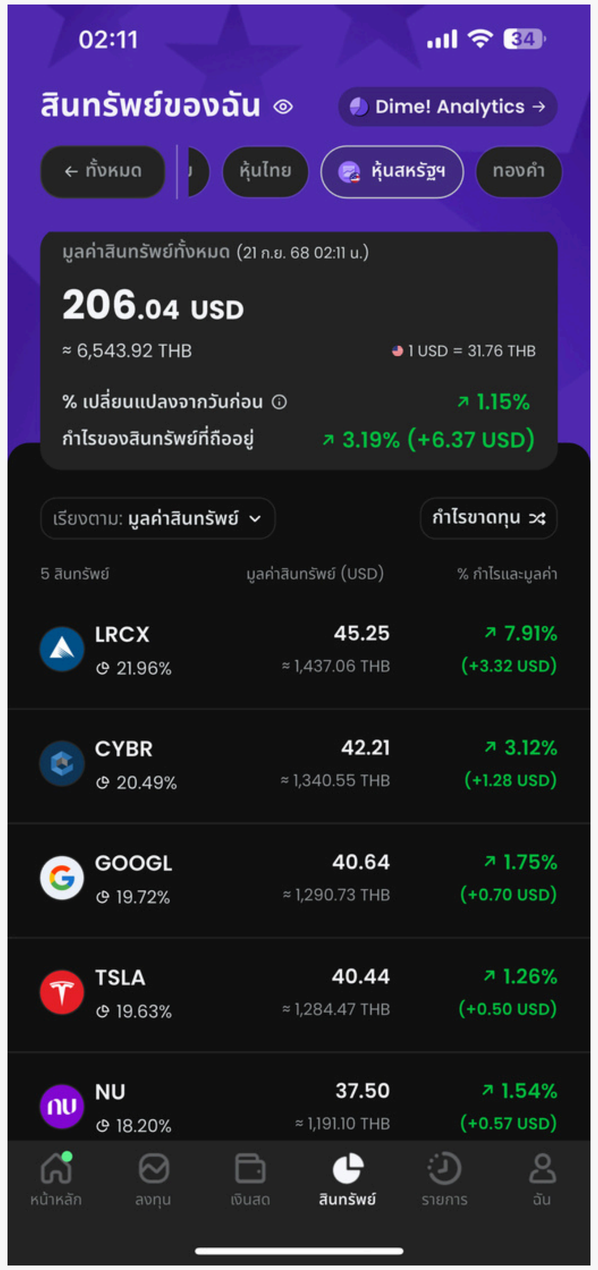
Example Portfolios from the 9-Week Experiment

6. Experimental Results

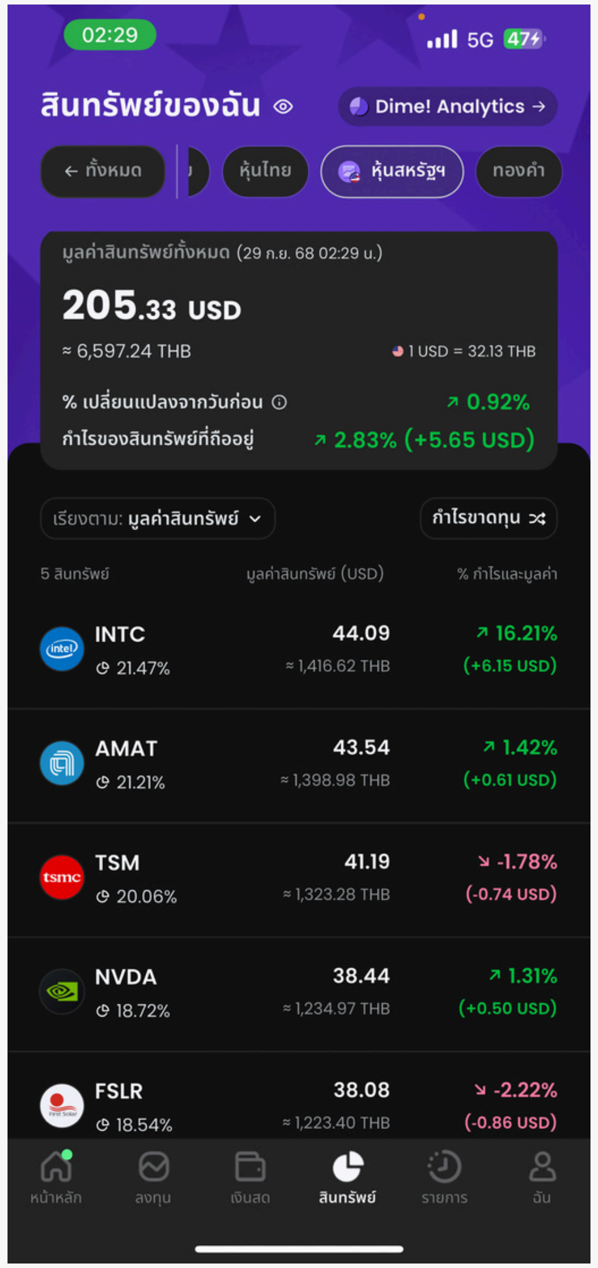
6.2 Live Investment Execution & Weekly Portfolios (2/2)



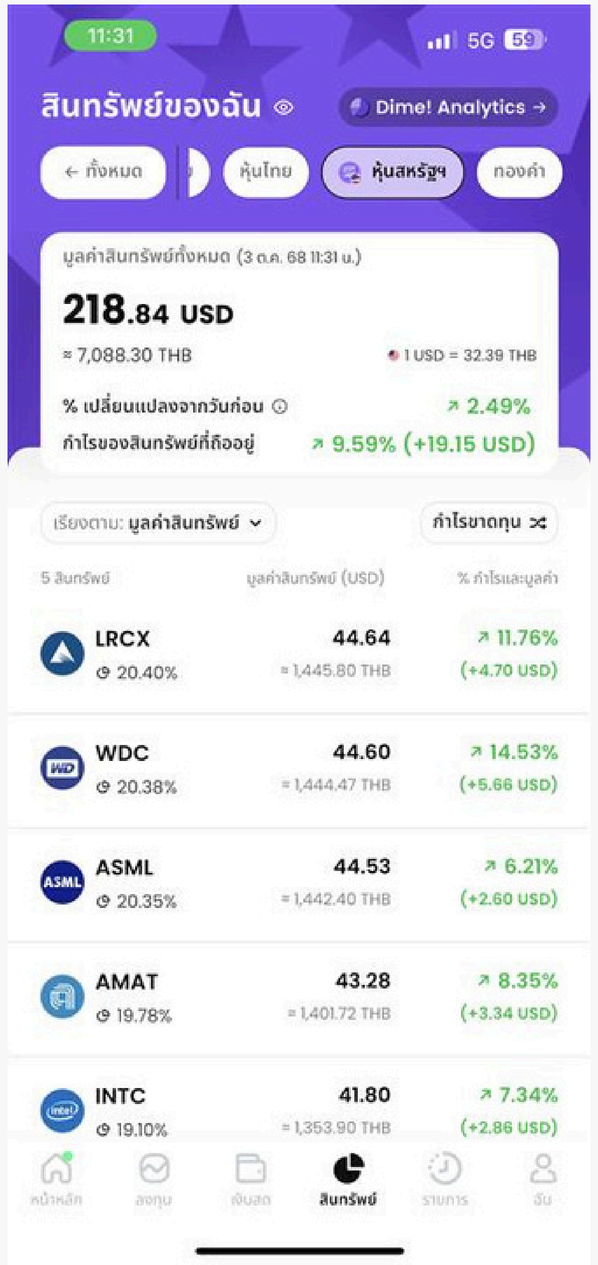
Sep 8-12, 2025



Sep 15-19, 2025



Sep 22-26, 2025



Sep 29 - Oct 3, 2025

Example Portfolios from the 9-Week Experiment

6.3 Performance Comparison

To evaluate our weekly strategy, we use the following metrics:

- **Average Weekly Return (%) :** Measures the mean percentage gain or loss per week.

$$\bar{r} = \frac{1}{n}\sum_{i=1}^{n} r_i$$

- **Cumulative Return (%) :** Represents the realized compounded portfolio growth over the entire period.

$$CR_{\mathrm{cum}} = \left(\prod_{i=1}^{n}\left(1 + \frac{r_i}{100}\right) - 1\right) \times 100$$

- **Standard Deviation (Weekly %) :** Measures volatility or dispersion of weekly returns.

$$\sigma_{\mathrm{wk}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(r_i - \bar{r})^2}$$

- **Sharpe Ratio (Annualized) :** Evaluates risk-adjusted return using weekly data.

$$Sharpe_{\mathrm{ann}} = \frac{\bar{r}_{\mathrm{wk}} - r_{f,\mathrm{wk}}}{\sigma_{\mathrm{wk}}} \times \sqrt{52}$$

- **Max Drawdown (Weekly %)** : Largest single-week decline in return during the period.

$$MDD_{\mathrm{weekly}} = \min(r_1, r_2, \ldots, r_n)$$

6.3 Performance Comparison

The table below details the week-by-week performance of the AI-selected portfolio against the NASDAQ-100, S&P 500, and the Tech Sector (XLK) ETF over the 9-week live investment period.

Table 3 . Weekly Performance Breakdown (AI Portfolio vs. Benchmarks)

Week	Date Range	AI Portfolio Return (%)	NASDAQ-100	S&P 500	Tech Sector (XLK)
1	Aug 4-8, 2025	**9.25%**	1.16%	1.17%	1.96%
2	Aug 11-15, 2025	**-1.40%**	0.30%	0.86%	-0.36%
3	Aug 18-22, 2025	**-1.23%**	-0.71%	0.36%	-1.16%
4	Aug 25-29, 2025	**1.27%**	-0.12%	0.06%	0.17%
5	Sep 2-5, 2025	**3.53%**	2.00%	1.26%	1.42%
6	Sep 8-12, 2025	**2.52%**	1.20%	1.37%	2.15%
7	Sep 15-19, 2025	**3.19%**	1.68%	0.71%	2.44%
8	Sep 22-26, 2025	**2.83%**	-0.52%	-0.26%	-0.01%
9	Sep 29 - Oct 3, 2025	**9.59%**	0.28%	0.67%	1.15%
Average Weekly Return		**3.28%**	**0.59%**	**0.69%**	**0.86%**

6.3 Performance Comparison

The table below summarizes the key performance metrics of the AI-selected portfolio against market benchmarks, calculated from the 9-week live investment experiment using a weekly strategy.

Table 4 . Performance Metrics Summary

Metric	AI Portfolio Return	NASDAQ-100	S&P 500	Tech Sector (XLK)
Cumulative Return	**32.9865%**	5.3562%	6.3607%	7.9662%
Average Weekly Return	**3.2833%**	0.5889%	0.6856%	0.8559%
Weekly Std Dev	**3.8601%**	0.9328%	**0.6385%**	1.2741%
Sharpe Ratio (Annualized)	**6.0232**	4.1166	**7.1044**	5.0291
Max Drawdown	**-1.40%**	-0.71 %	**-0.26 %**	-1.16 %

Figure 1: Weekly Return Comparison vs. Benchmarks

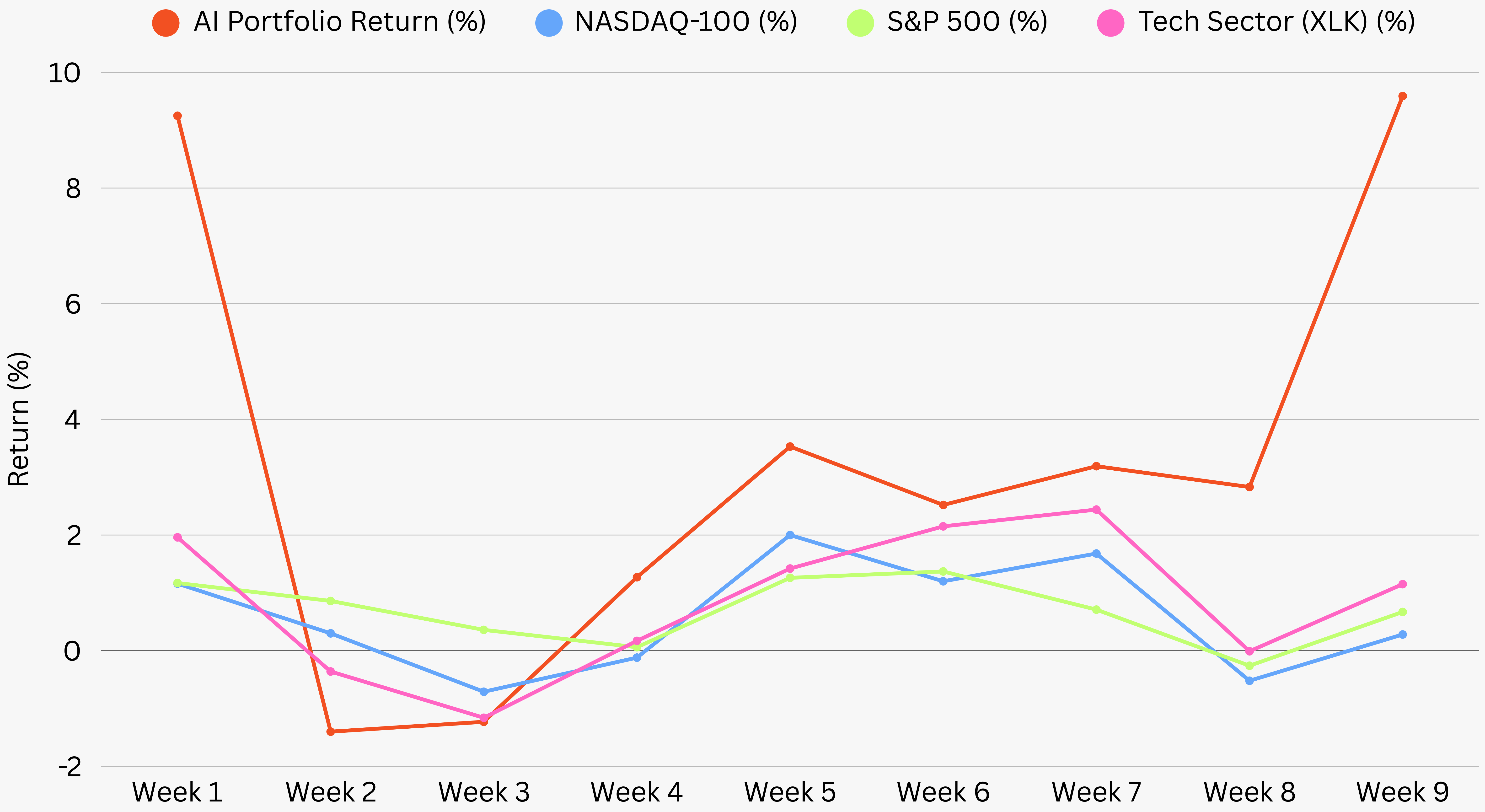
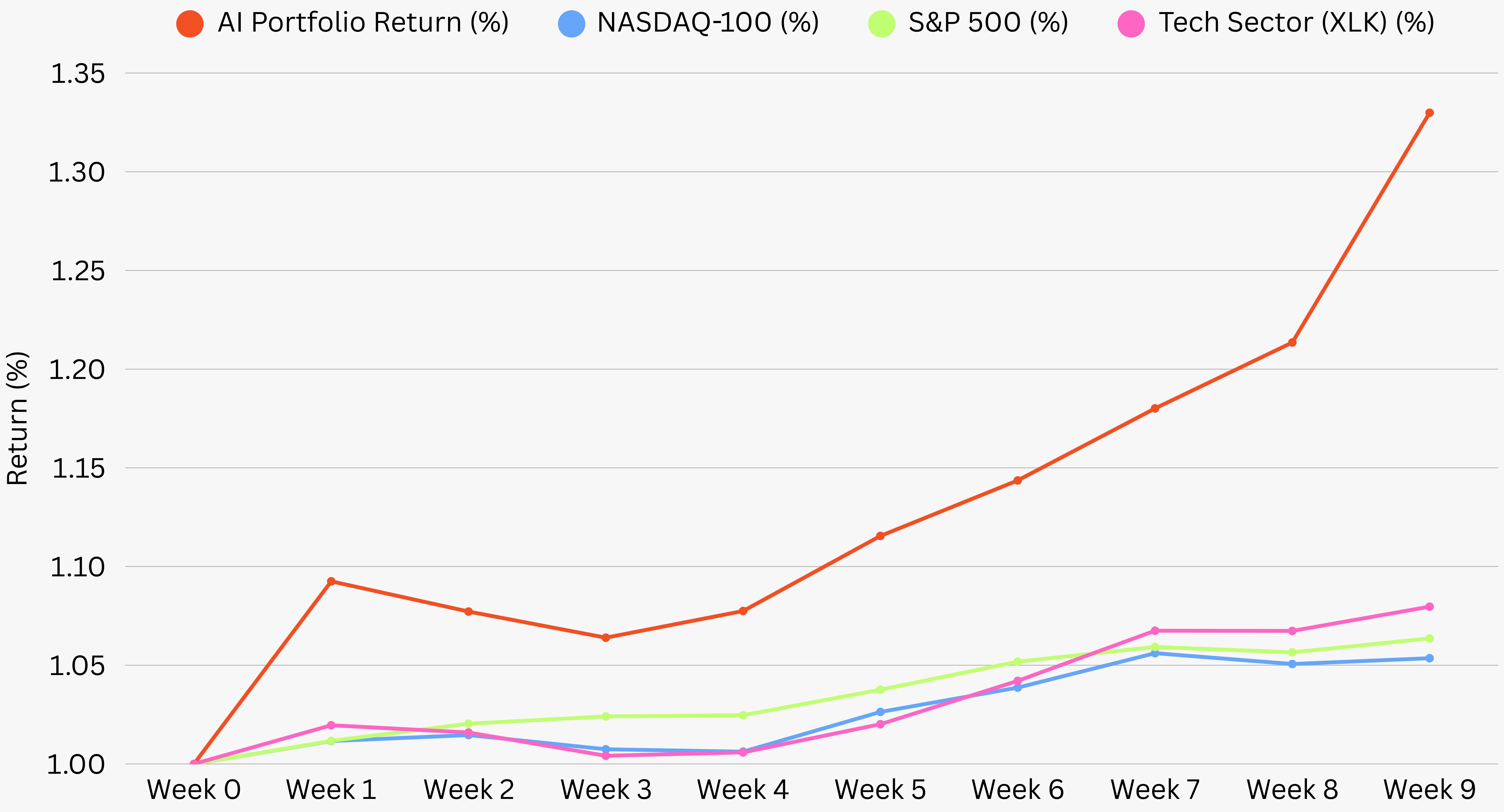


Figure 2: Theoretical Cumulative Return vs. Benchmarks



6. Experimental Results

6.4 Discussion (1/2)

The 9-week live investment experiment confirms the Agentic AI system's ability to achieve superior short-term performance compared to benchmarks.

- Consistent Profitability: Achieved +3.28% Average Weekly Return, significantly exceeding benchmarks (+0.59% to +0.86%), validating the core strategy effectiveness per cycle.
- Strong Growth Potential: Delivered +33.0% Theoretical Cumulative Return, vastly outperforming indices (NASDAQ: +5.4%, S&P 500: +6.4%, XLK: +8.0%), highlighting potent alpha generation capability if compounded.
- Efficient Risk-Adjusted Returns: Despite higher volatility (Std Dev: 3.86%), the Annualized Sharpe Ratio of 6.02 demonstrates excellent return generation per unit of weekly risk, surpassing tech benchmarks.
- Controlled Downside: A Max Drawdown (Weekly) of only -1.40% indicates effective capital preservation within single trading cycles, mitigating downside risk effectively.

6. Experimental Results

6.4 Discussion (2/2)

Limitations of This Study

- This study employed only one large language model — Llama 3 70B running on Groq — without comparison to other foundation models such as GPT-4 or Claude 3.
- The system relied solely on zero-shot reasoning, without applying any domain-specific fine-tuning, which may limit its understanding of financial and technical terminology.
- The live experiment covered only nine weeks and focused exclusively on the U.S. Tech sector, limiting generalization to other industries or longer timeframes.
- The current framework lacks an adaptive learning or feedback loop, requiring manual parameter adjustments between weekly execution cycles.

7. Conclusions

7.1 Key Findings & Contribution

This research successfully developed and validated a specialized Agentic AI framework via live investment experiment, demonstrating a significant performance edge in short-term tech stock selection. The system consistently generated alpha, achieving a +3.28% average weekly return ($59.10 total profit on $200 weekly capital) with a 77.8% win rate, decisively outperforming market benchmarks over 9 weeks.

Critically, the high returns were achieved with efficient risk management, evidenced by a strong Annualized Sharpe Ratio of 6.02 and a contained Max Drawdown (Weekly) of -1.40%. These results validate the effectiveness of the 5-specialist-agent architecture (Momentum, Catalyst, Technical, Liquidity, Risk) orchestrated by LangGraph. A key contribution lies in the system's explainability, derived from agent-specific scoring, offering transparency beyond typical black-box models and proving the practical viability of this agentic approach in real-world conditions.

7. Conclusions

7.2 Future Work

While the live investment results are promising, this research is a foundational step with clear paths for future enhancement. Key limitations include API rate constraints and the 9-week validation period.

Future Work & Enhancements:

- Future Extension: Enable inter-agent communication or debate mechanisms inspired by TradingAgents (Xiao et al., 2025)., to enhance decision robustness.
- Extending Investment Horizons: Adapt agent logic and scoring models to support longer-term strategies (1-month, quarterly, multi-year holds) to capture different market dynamics.
- End-to-End Automation: Integrate with brokerage APIs for fully automated, real-time investment execution, removing manual steps.
- Dynamic Agent Weighting & Self-Correction: Develop a "meta-agent" that adjusts agent score weights based on market regimes (e.g., increasing Risk Agent's influence during volatility) and implements feedback loops for agents to learn from past performance.

References (1/4) (APA reference)

Agentic AI & Multi-Agent Financial Systems

[1] Xiao, Y., Sun, E., Luo, D., & Wang, W. (2024). TradingAgents: Multi-agents LLM financial trading framework. arXiv preprint arXiv:2412.20138.

[2] Feng, J., Du, Y., Zhao, J., & Li, Y. (2024). Agentmove: A large language model based agentic framework for zero-shot next location prediction. arXiv preprint arXiv:2408.13986.

[3] Chen, W., Su, Y., Zuo, J., Yang, C., Yuan, C., Qian, C., ... & Zhou, J. (2023). Agentverse: Facilitating multi-agent collaboration and exploring emergent behaviors in agents. arXiv preprint arXiv:2308.10848, 2(4), 6.

[4] Ding, H., Li, Y., Wang, J., & Chen, H. (2024). Large language model agent in financial trading: A survey. arXiv preprint arXiv:2408.06361.

[5] Hong, S., Zheng, X., Chen, J., Cheng, Y., Wang, J., Zhang, C., ... & Wu, C. (2023). MetaGPT: Meta Programming for Multi-Agent Collaborative Framework. CoRR abs/2308.00352 (2023). URL: https://doi. org/10.48550/arXiv, 2308.

[6] Zhu, K., Wang, J., Zhao, Q., Xu, R., & Xie, X. (2024). Dynamic evaluation of large language models by meta probing agents. arXiv preprint arXiv:2402.14865.

[7] Saturwar, J., Ghetiya, M., Kolipaka, N., Maurya, V., & Maurya, P. (2025). StockVision: A Multi-Agent AI framework for Intelligent Stock Investment and Analysis using CrewAI. ITEGAM-JETIA, 11(55), 01-10.

[8] Patel, N., & Chauhan, H. (2025). Agentic AI and the Future of Work: Transforming Labor Markets, Economic Structures, and Workforce Development. In The Power of Agentic AI: Redefining Human Life and Decision-Making: In Industry 6.0 (pp. 205-227). Cham: Springer Nature Switzerland.

References (2/4) (APA reference)

LLMs in Financial Analysis & Trading

[9] Fatouros, G., Metaxas, K., Soldatos, J., & Kyriazis, D. (2024). Can large language models beat wall street? unveiling the potential of ai in stock selection. arXiv preprint arXiv:2401.03737.

[10] Lopez-Lira, A. (2024). Can ChatGPT forecast stock price movements?.

[11] Henning, T., Ojha, S. M., Spoon, R., Han, J., & Camerer, C. F. (2025). LLM Trading: Analysis of LLM Agent Behavior in Experimental Asset Markets. arXiv preprint arXiv:2502.15800.

[12] Xie, Q., Han, W., Zhang, X., Lai, Y., Peng, M., Lopez-Lira, A., & Huang, J. (2023). Pixiu: A large language model, instruction data and evaluation benchmark for finance. arXiv preprint arXiv:2306.05443.

[13] Yang, H., Liu, X., & Wang, C. D. (2023). Fingpt: Open-source financial large language models. CoRR, abs/2306.06031, 2023. doi: 10.48550. arXiv preprint ARXIV.2306.06031, 1.

[14] Ma, T., Du, J., Huang, W., Wang, W., Xie, L., Zhong, X., & Zhou, J. T. (2025). Llm knows geometry better than algebra: Numerical understanding of llm-based agents in a trading arena. arXiv preprint arXiv:2502.17967.

[15] Darwish, M., Hassanien, E. E., & Eissa, A. H. (2025). Stock Market Forecasting: From Traditional Predictive Models to Large Language Models: M. Darwish et al. Computational Economics, 1-45.

[16] Papasotiriou, K., Sood, S., Reynolds, S., & Balch, T. (2024, November). Ai in investment analysis: Llms for equity stock ratings. In Proceedings of the 5th ACM International Conference on AI in Finance (pp. 419-427).

[17] Wu, S., Irsoy, O., Lu, S., Dabravolski, V., Dredze, M., Gehrmann, S., ... & Mann, G. (2023). Bloomberggpt: A large language model for finance. arXiv preprint arXiv:2303.17564.

References (3/4) (APA reference)

Orchestration, Prompting & Explainability

[18] LangChain, A. I. (2024). Langgraph: Building structured applications with llms.

[19] Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2023). ReAct: Synergizing reasoning and acting in language models (arXiv: 2210.03629). arXiv.

[20] White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., ... & Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with chatgpt. arXiv preprint arXiv:2302.11382.

[21] Xu, B., Yang, A., Lin, J., Wang, Q., Zhou, C., Zhang, Y., & Mao, Z. (2023). Expertprompting: Instructing large language models to be distinguished experts. arXiv preprint arXiv:2305.14688.

[22] Shinn, N., Labash, B., & Gopinath, A. Reflexion: an autonomous agent with dynamic memory and self-reflection. CoRR, abs/2303.11366, 2023. doi: 10.48550. arXiv preprint ARXIV.2303.11366.

[23] Danilevsky, M., Qian, K., Aharonov, R., Katsis, Y., Kawas, B., & Sen, P. (2020). A survey of the state of explainable AI for natural language processing. arXiv preprint arXiv:2010.00711.

[24] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. ACM computing surveys, 55(12), 1-38.

[25] Roumeliotis, K. I., Sapkota, R., Karkee, M., & Tselikas, N. D. (2025). Agentic AI with Orchestrator-Agent Trust: A Modular Visual Classification Framework with Trust-Aware Orchestration and RAG-Based Reasoning. arXiv preprint arXiv:2507.10571.

[26] Sapkota, R., Roumeliotis, K. I., & Karkee, M. (2025). Ai agents vs. agentic ai: A conceptual taxonomy, applications and challenges. arXiv preprint arXiv:2505.10468.

References (4/4) (APA reference)

Performance Evaluation & Benchmarking & Applications

[27] Li, W. W., Kim, H., Cucuringu, M., & Ma, T. (2025). Can LLM-based Financial Investing Strategies Outperform the Market in Long Run?. arXiv preprint arXiv:2505.07078.

[28] Koa, K. J., Ma, Y., Ng, R., & Chua, T. S. (2024, May). Learning to generate explainable stock predictions using self-reflective large language models. In Proceedings of the ACM Web Conference 2024 (pp. 4304-4315).

[29] Li, Y., Yu, Y., Li, H., Chen, Z., & Khashanah, K. (2023). Tradinggpt: Multi-agent system with layered memory and distinct characters for enhanced financial trading performance. arXiv preprint arXiv:2309.03736.

[30] Kang, H., Zhang, Q., Cai, H., Xu, W., Krishna, T., Du, Y., & Weissman, T. (2025). Win Fast or Lose Slow: Balancing Speed and Accuracy in Latency-Sensitive Decisions of LLMs. arXiv preprint arXiv:2505.19481.

[31] Sharpe, W. F. (1966). Mutual fund performance. The Journal of business, 39(1), 119-138.

[32] Pereira, P. L., Caldeira, J., & Sebastião, H. Artificial Intelligence Driven Portfolio Recommendation System: Leveraging Sentiment Analysis Towards Improved Performance. Available at SSRN 5000504.

[33] Jensen, M. C. (1968). The performance of mutual funds in the period 1945-1964. The Journal of finance, 23(2), 389-416.

[34] Joshi, S. (2025). Comprehensive Review of Artificial General Intelligence AGI and Agentic GenAI: Applications in Business and Finance. Available at SSRN 5250611.

[35] Kirkovska, A. (2025). Llama 3 70B vs GPT-4: Comparison Analysis. Vellum AI Reports. Retrieved from https://www.vellum.ai/blog/llama-3-70b-vs-gpt-4-comparison-analysis

Thank you

Questions & Discussion

Advisor: Asst. Prof. Dr. Ekarat Rattagan
Advisee: Kunakorn Pruksakorn 6610422020
Graduate School of Applied Statistics (DADS5), NIDA