

Language Model Guided Reinforcement Learning in Quantitative Trading

LLM-Guided Intelligence Bridging Strategy and Safety in Trading

Adam Darmanin

Supervised by Dr. Vince Vella

Department of AI
Faculty of ICT
University of Malta

November, 2025

*A thesis submitted in partial fulfilment of the requirements for the degree
of M.Sc. in AI.*



L-Università
ta' Malta

University of Malta Library – Electronic Thesis & Dissertations (ETD) Repository

The copyright of this thesis/dissertation belongs to the author. The author's rights in respect of this work are as defined by the Copyright Act (Chapter 415) of the Laws of Malta or as modified by any successive legislation.

Users may access this full-text thesis/dissertation and can make use of the information contained in accordance with the Copyright Act provided that the author must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the prior permission of the copyright holder.



**L-Università
ta' Malta**

Copyright ©2025 University of Malta

WWW.UM.EDU.MT

First edition, November 17, 2025

Acknowledgements

I would like to express my gratitude to Dr. Vince Vella for his supervision and guidance. His support was instrumental at every stage of this research.

I am also thankful to the expert reviewers who generously contributed their time, insight, and domain expertise to this project. Their feedback played a critical role in shaping the strategies generated by the Large Language Model (LLM) agents. I am particularly grateful to Yuri Perin (CFO, Soteria Shielding System; former Quant at BlackRock), Anthony Roberts (CEO, KnowMio; former Managing Director at BlackRock), Apostolia Tasoudi (Lead Financial Engineer, BlackRock), Albert Viilik (Quantitative Analyst, BlackRock), Alex De Castro, PhD (Lead Researcher, BlackRock), Admir Soko (Director, BlackRock), Chris Tubby (Expert Trading Coach), Andre Darmanin (Investment Advisory Board Member, Aramis Capital), as well as Aitor Souto and Nathan Ciantar, both independent retail traders. Their diverse perspectives enriched the outcome of the LLM agents.

Abstract

This research explores the application of LLMs in guiding Reinforcement Learning (RL) algorithms to address key challenges in algorithmic trading. While RL agents are effective at optimizing actions based on reward signals, they often exhibit myopic behavior, lacking the strategic foresight and economic intuition needed to operate in a complex environment such as the financial markets.

To address this limitation, LLMs are introduced as strategic planners capable of synthesizing high-level trading strategies from heterogeneous sources, including market data, macroeconomic indicators, and news sentiment. By informing the policy layer of RL agents, LLMs enable the generation of trading strategies that are both risk-aware and sensitive to the prevailing market conditions. Unlike traditional approaches that retrain or control the RL agent per scenario, the LLM acts as a guidance mechanism that adapts its outputs to align with predefined high-risk and low-risk investor profiles, thereby enabling the same underlying RL agent to operate effectively across distinct risk preferences.

We propose a novel framework to evaluate the potential of LLMs guidance for RL agents. The research has two objectives: (i) to determine whether LLMs can reliably produce market-aware strategies that meet the standards of professional trading systems, as validated through Human-in-the-Loop (HITL) expert surveys; and (ii) to assess whether LLMs can guide a single RL agent to improve its trading performance, measured with the Sharpe Ratio (SR) in a high-risk setting, and enhance its risk management, measured through the Maximum Drawdown (MDD) in a low-risk setting, without retraining or modifying the agent itself.

To achieve these objectives, LLMs generate high-level strategies using prompt templates iteratively refined through expert-provided trades and validated through back-testing using the metric SR. The performance of the LLM-guided Deep Reinforcement Learning (DRL) agent is benchmarked against a traditional RL model using a subset of securities within the technology sector. The final RL agent's effectiveness is assessed using standard portfolio metrics, including SR and MDD. Empirical results demonstrate that LLM-enhanced RL agents can achieve superior SR and MDD than their benchmarks, with greater safety when compared to black-box deep learning counterparts.

This research contributes to the fields of algorithmic trading and Artificial Intelligence (AI) by demonstrating the synergistic application of LLMs to RL, highlighting their ability to adapt a single learning agent to diverse investor risk profiles and enhance performance and safety in complex decision-making environments.

Contents

List of Abbreviations	x
1 Introduction	1
1.1 Algorithmic Trading	1
1.2 Motivation	3
1.3 Aim and Objectives	4
1.4 Proposed Solution	5
1.5 Contributions	5
1.6 Document Structure	6
2 Background & Literature Overview	7
2.1 Algorithmic Trading	7
2.1.1 Long-term Investment	7
2.1.2 Short-term Trading	8
2.1.3 Investor Risk Profiles	13
2.2 Reinforcement Learning in Trading	13
2.3 LLMs in Decision-Making	20
2.4 Long-Term Planning Challenges in RL	24
2.5 Related Works	25
2.5.1 Deep Reinforcement Learning (DRL)	25
2.5.2 Large Language Model (LLM)	28
2.5.3 Large Language Model (LLM) and Reinforcement Learning (RL)	32
2.5.4 Investor Risk Profiles	34
2.6 Conclusion	35
3 Materials & Methods	36
3.1 Benchmark Research	37
3.2 Experiments	38

3.2.1	Experiments Overview	38
3.2.2	Experiment 1	38
3.2.3	Experiment 2	58
3.3	Conclusion	62
4	Results & Discussion	64
4.1	Experiment 1 Results	64
4.1.1	Discussion	65
4.2	Experiment 2a Results - High Risk Profile	73
4.2.1	Discussion: Experiment 2a (High Risk)	73
4.3	Experiment 2b Results (Low Risk)	75
4.3.1	Discussion: Experiment 2b (Low Risk)	77
4.4	Overall Findings	78
5	Conclusion	79
5.1	Revisiting Aims and Objectives	79
5.2	Critique and Limitations	80
5.3	Future Work	81
5.4	Final Remarks	82
	References	83
	Appendix A Algorithms	92
	Appendix B Equations	96
B.1	Financial and Technical Formulas	96
B.2	Technical Indicators	98
B.3	Volatility and Derivatives-Based Metrics	98
	Appendix C Prompts	100
C.1	Baseline Strategy Prompt	100
C.2	Writer-Trainer Prompt	102
C.3	Writer-Generator Prompt	108
C.4	Judge Prompt v1	110
C.5	Strategy Prompt v1: Tuned Prompt	111
C.6	Strategy Prompt v2: In-Context Memory Prompt	114
C.7	Strategy Prompt v3: Decomposed Prompting	118
C.8	Strategy Prompt v4: News Factors	123
C.9	News Factors Prompt	128

Appendix D Data	130
D.1 Replicated Benchmark Metrics	130
D.2 Data Availability	130
Appendix E Replication and Reproducibility	132
E.1 Software Environment	132
E.2 Hardware Configuration	133
E.3 Random Seeds and Replications	133
E.4 Data Processing	133
E.5 Prompt and Model Configuration	134
Appendix F Expert Reviews	135

List of Figures

2.1	Stock Moves Visualization	9
2.2	AI Techniques utilized Trading (Dakalbab et al., 2024)	12
3.1	Prompt Tuning Workflow	47
3.2	Prompt Refinement Loop with Writer-Judge (S. Wang et al., 2024)	49
3.3	LLM Agents Architecture.	54
3.4	LLM/RL Hybrid Architecture.	61
4.1	Top quartile features used under high-risk configuration.	68
4.2	Top quartile bigrams in explanations under high-risk configuration	68
4.3	Top quartile features used under low-risk configuration	69
4.4	Top quartile bigrams in explanations under low-risk configuration	70
4.5	AAPL Performance with LLM+RL Model.	75
4.6	Training Behavior for AAPL: Q-Values for LONG.	76
4.7	Training Behavior for AAPL: Q-Values for SHORT.	76
4.8	Training Behavior for AAPL: Sharpe Ratio.	77

List of Tables

2.1	Evaluation of LLM Performance in Financial Research (2022–2025).	22
3.1	Knowledge Cutoff Dates for Generative Pretrained Transformer (GPT) (OpenAI, 2024)	46
3.2	Rubric for evaluating LLM-generated trading rationales.	58
3.3	Benchmark thresholds.	62
4.1	Prompt Versions for Experiment 1	64
4.2	Sharpe Ratio (SR) per Prompt and Ticker under Risk and No-Risk Profiles.	65
4.3	Perplexity (PPL) per Prompt and Ticker for Risk Profiles.	65
4.4	Entropy (H_{LLM}) per Prompt and Ticker for Risk Profiles.	65
4.5	t -test p -values for Metric Changes Between P1 and P4 under Risk and No-Risk Profiles.	66
4.6	Expert Reviewer Scores (ERS) for the final Prompt 4 (1–3 rubric).	66
4.7	Experiment 2a. Sharpe Ratio (SR) for High-Risk Profile. <i>Note.</i> Paired t -tests compare LLM+RL against RL-Only.	73
4.8	Experiment 2b. Maximum Drawdown (MDD) for Low-Risk Profile.	77
D.1	Replication Metrics for Théate and Ernst (2021)	131
E.1	Core software versions used in all experiments	132
F.1	Expert Review Participant Demographics (Aggregated Summary)	135

List of Abbreviations

DDQN	Double Deep Q-Network	16
LLM	Large Language Model	2
CoT	Chain of Thought	51
AI	Artificial Intelligence	3
ML	Machine Learning	1
RL	Reinforcement Learning	1
DRL	Deep Reinforcement Learning	7
HRL	Hierarchical Reinforcement Learning	24
PPO	Proximal Policy Optimization	19
DQN	Deep Q-Network	15
RNN	Recurrent Neural Network	20
GPT	Generative Pretrained Transformer	21
ICL	In-Context Learning	23
MDP	Markov Decision Process	14
HITL	Human-in-the-Loop	48
ICM	In-Context Memory	51
OHLCV	Open, High, Low, Close and Volume	41
IV	Implied Volatility	39
RSI	Relative Strength Index	47
MACD	Moving Average Convergence Divergence	47
SR	Sharpe Ratio	3
MDD	Maximum Drawdown	5
MDD_{dur}	Maximum Drawdown Duration	38
PPL	Perplexity	47
SMA	Simple Moving Average	47
PM	Portfolio Manager	12
LOB	Limit Order Book	80
EMH	Efficient Markets Hypothesis	12
HFT	High-Frequency Trading	7
A2C	Advantage Actor Critic	18
A3C	Asynchronous Advantage Actor Critic	18
SAC	Soft Actor Critic	19
ERS	Expert Review Score	64

Introduction

1.1 | Algorithmic Trading

Algorithmic trading involves the development of autonomous systems that execute trades by leveraging statistical, econometric, and, in some cases, Machine Learning (ML) techniques. These systems process large datasets of structured and unstructured data across various frequencies and respond to market dynamics in real-time. A primary challenge in this domain is the design of models that make rational, human-understandable decisions while maintaining robust performance in inherently complex market environments (Liu, Yang, et al., 2022; Pricope, 2021).

However, the practical implementation of these models becomes significantly more complex when applied to the structured workflows of professional investment domains. In these environments, trade ideas are not merely generated algorithmically but are developed through a rigorous hierarchy: macroeconomic research informs initial concepts, which are refined by specialized teams, reviewed by expert committees, and ultimately executed by trading desks. This structured approach ensures alignment with organizational objectives and shareholder interests, creating a robust decision-making framework for deploying capital.

Integrating machine learning models into such workflows introduces additional challenges. One being the critical requirement to ensuring the algorithm's safety to the trading domain, as modern ML algorithms, while powerful, often detect patterns that may be spurious or overfitted (López de Prado, 2018a, 2020b). These limitations make it difficult to replicate the nuanced reasoning and oversight provided by human experts in high stakes trading environments.

Algorithmic trading systems have evolved significantly, transitioning from early rule-based strategies to advanced machine learning based approaches.

Of these advanced ML techniques, Reinforcement Learning (RL) has emerged as a dominant paradigm, garnering considerable enthusiasm from both retail and professional traders due to its demonstrated ability to perform effectively in most stochastic

environments, financial markets being one (Liu, Yang, et al., 2022; Pricope, 2021; Xu et al., 2023). RL has enabled the development of intelligent, autonomous trading agents capable of learning from and reacting to price movements, technical indicators, and economic signals. These agents frame trading as a sequential decision-making task, approximating the future value of their observations to derive actionable insights from vast datasets. RL models have demonstrated effectiveness in trading specific assets within constrained market regimes (Corazza & Bertoluzzo, 2014; Takara et al., 2023; Théate & Ernst, 2021), with Q-learning, Soft-Actor-Critic and Proximal Policy Optimization having proven performance in the stock market (Liu, Yang, et al., 2022; Xu et al., 2023).

Despite these advancements, RL models face significant challenges. Issues such as overfitting during training (Devidze et al., 2022; X. Wang et al., 2020), inefficient exploration-exploitation strategies, and poorly aligned reward functions often result in policies that are unsafe, inflexible, or detached from real-world trading objectives (Booth et al., 2023; Pricope, 2021). Furthermore, their lack alignment with the reasoning processes of human traders undermines trust in their outputs (Yu et al., 2024). There is, therefore, a need for guiding systems that balance performance, safety, and fidelity to professional practices. Specifically, the decisions made by the guided models should exhibit clear economic intuition to ensure their alignment with real-world trading objectives (López de Prado, 2018a, 2020b).

Large Language Model (LLM) have recently demonstrated exceptional capabilities in complex reasoning, and environmental awareness (A. H. Huang et al., 2023; Mezghani et al., 2023; H. Yang et al., 2023). Within financial contexts, LLMs have shown promise in analyzing high-dimensional datasets, generating trade signals based on news insights, and processing market sentiment (Lopez-Lira & Tang, 2023; Onozo et al., 2024; Yu et al., 2024; Zhou et al., 2021). However, while LLMs demonstrate significant potential, their applications in algorithmic trading have produced mixed results, with challenges including sub-optimal decision-making, risks of confabulation, and limitations in handling numerical data or temporal reasoning, as their strengths primarily lie in natural language processing (M. Wang et al., 2024; Zhang et al., 2024). Some of the literature address these limitations by verbalizing numerical data points into symbolic tokens, enabling LLMs to process them more effectively (Hu et al., 2024; Schulhoff et al., 2024; Yu et al., 2024).

Drawing inspiration from the hierarchical decision-making models used in professional investment workflows, this research proposes a novel framework that integrates LLMs with RL agents to emulate a multi-layered paradigm for algorithmic trading. In the proposed framework, LLMs function as strategy generators, synthesizing trade ideas from diverse data sources, and guiding the policies of RL agents through strategies

vetted for alignment with professional heuristics. This integration enhances the safety of the algorithm, and fidelity of its trading policies to the structured decision-making processes employed by professionals.

1.2 | Motivation

The motivation for this research stems from the growing complexity of algorithmic trading systems and the challenges associated with their adoption in professional financial environments. Modern trading systems must process vast volumes of heterogeneous data and respond to rapidly evolving market conditions. The scale of information, the velocity of trading, and the available compute have created unprecedented opportunities for Artificial Intelligence (AI); yet, much of this potential remains unrealized due to a persistent lack of trust in the outputs of learning-based systems.

Despite notable progress in deep learning, these methods have yet to gain widespread traction in institutional finance. Concerns about reliability, auditability, and regulatory compliance persist. Financial markets demand transparency and accountability, with legislative frameworks such as MiFID II (Parliament & of the European Union, 2014), the EU AI Act (Commission, 2021), and GDPR Article 22 (Bygrave, 2020) placing stringent constraints on the deployment of opaque or non-interpretable models. A black-box system that cannot explain its decisions not only fails to inspire confidence, it also introduces legal and reputational risk.

Among deep learning methods, RL faces particular skepticism. RL agents, typically trained through trial-and-error within narrow environments, struggle to replicate the economic intuition of professional human traders. Their uninterpretable policies have limited their adoption, despite the publication of superficially promising Sharpe Ratio (SR) (Bailey & López de Prado, 2014).

LLMs represent a new generation of AI capable of synthesizing complex information across domains. Their capacity to interpret natural language and reason over unstructured financial data enables them to emulate the high-level decision-making observed in discretionary strategies (Onozo et al., 2024; Zhang et al., 2024). Building on this capacity, this research investigates whether LLM agents can guide a RL agent to produce coherent, economically grounded strategies aligned with investor risk tolerances.

Rather than changing the RL architecture for each risk configuration, the LLM can modulate the behavior of the RL policy through strategic prompting. This enables the same underlying agent to be adapted dynamically to diverse investor risk profiles solely on changes in LLM-generated guidance. The LLM will not act as a controller, but as a

strategic planner capable of aligning RL agents with investor-specific risk configurations.

LLMs bring their own challenges, with confabulation being among the most significant. Confabulation refers to the generation of outputs that, while syntactically coherent, are factually incorrect or fabricated. This risk arises from the probabilistic nature of LLMs, which prioritize fluency and contextual relevance over factual accuracy (Gonen et al., 2023; Schulhoff et al., 2024). Structured prompting methods such as “Self-Consistency” and “Chain-of-Verification” (Schulhoff et al., 2024) attempt to mitigate these effects by encouraging logical reasoning paths.

The problem is inherently non-trivial. Trust is not a technical feature but a prerequisite for the adoption of any AI in high-stakes domains like finance. State-of-the-art models are often underutilized because professionals remain skeptical of outputs they cannot understand or trace, fearing catastrophic errors or overreliance on incomprehensible patterns (López de Prado, 2018a, 2018b). By combining the contextual reasoning of LLMs with RL’s tactical capabilities, this research aims to build a framework that bridges the gap between performance and trust.

1.3 | Aim and Objectives

The aim of this research is to investigate the integration of LLM agents with RL agents for algorithmic trading. To support this aim, the following objectives are articulated as testable research questions, each corresponding to a core component of the proposed hybrid framework:

- **Objective 1: LLM Trading Strategy Generation** - Can an LLM generate trading strategies that are coherent, economically grounded, and aligned with predefined risk constraints, as assessed by expert reviewers?
- **Objective 2: LLM-Guided RL Agent Performance Across Risk Profiles** - To evaluate whether integrating LLM-generated strategies into RL agents improves portfolio performance relative to benchmark thresholds under varying risk settings. All agents share the same underlying architecture, with risk alignment driven solely by the LLM guidance.
 - **Objective 2a (High-Risk):** Does LLM-guided agent integration improve the SR relative to unguided baselines in high-risk trading scenarios?

- **Objective 2b (Low-Risk):** Does LLM-guided agent reduce Maximum Draw-down (MDD) compared to unguided baselines in low-risk trading environments?

1.4 | Proposed Solution

The proposed solution emulates the industry’s structured decision-making workflows by integrating LLMs with RL agents in a hierarchical architecture (Zhang et al., 2024). It employs three distinct agents: a Strategist Agent ($\mathcal{A}_{\text{strategist}}$), an Analyst Agent $\mathcal{A}_{\text{analyst}}$, and a tactical RL Agent \mathcal{A}_{rl} . These agents collaborate to process diverse market signals, generate trading strategies, and guide the RL to execute the decisions, at the best possible price and time.

The workflow begins by processing a diverse set of market signals, including macroeconomic indicators, fundamental ratios, price and technical metrics, volatility measures, and market news. The Strategist Agent synthesizes these signals to generate high-level global policies (π^g) that outline potential trading strategies. The Analyst Agent transforms news into actionable factors and integrates them into these strategies (Onozo et al., 2024; M. Wang et al., 2024), enhancing contextual awareness.

The global policies (π^g) are passed to the RL Agent, which adapts them into decisions through its local policy (π^l). The RL Agent’s local policy ensures responsiveness to real-time environmental states while maintaining alignment with global strategic objectives and preserving a level of safety and performance guided by the LLM. Crucially, the same underlying RL agent is reused across different investor risk profiles, with the LLM adapting its strategy generation to reflect profile-specific constraints and preferences.

The entire workflow mirrors the structured decision-making processes of the financial industry, ensuring a degree of fidelity with professional standards from the outset.

1.5 | Contributions

This research advances the domain of algorithmic trading through the development of a novel framework that integrates LLMs with RL agents.

Firstly, this work proposes a methodology for leveraging LLMs to generate trading strategies. By employing carefully designed prompts and techniques to create these, LLMs are guided to produce strategies that have economic rationale. These strategies

are validated by domain experts to ensure fidelity to professional heuristics and practical applicability.

Secondly, the research demonstrates the application of LLM-generated strategies to guide RL agents. The integration mechanisms enable RL agents, which traditionally operate with limited domain knowledge and strategic capabilities beyond their observation space, to adapt effectively to real-world environments where strategic decision-making and contextual understanding are essential. Importantly, the same RL agent is deployed across two risk configurations, with the LLM modulating its strategic guidance according to the risk. This allows for modular adaptability to distinct risk regimes without retraining the agent itself.

This research advances the understanding of the synergy between reasoning and decision-making in AI by demonstrating how LLMs can guide RL agents to emulate the structured, hierarchical workflows used in professional trading environments. These findings extend beyond algorithmic trading, providing valuable insights into models applied in complex, real-world decision-making frameworks.

1.6 | Document Structure

The remainder of this document is structured as follows:

Chapter 2 establishes the foundational knowledge required for the work presented in this study. It provides an overview of LLMs, reinforcement learning (RL), and quantitative trading concepts. Chapter 2 also delivers a review of prior research related to the integration of LLMs with RL.

Chapter 3 details the proposed methodology for integrating LLM-generated strategies with RL agents, introducing the integration mechanisms, rationale engineering techniques, and the evaluation frameworks to assess the results.

Chapter 4 presents the experimental results, assessing the performance of the proposed integration mechanisms in handling real-world constraints and discusses their implications, highlighting the strengths, limitations, and potential improvements of the proposed solution within the broader field of RL and LLM integration.

Chapter 5 concludes by revisiting the key contributions of this work, summarizes the outcomes, and identifies directions for future research.

Appendix A includes all algorithms used, Appendix B presents the equations used for in feature engineering, Appendix C includes all prompts used in this study, with Appendix D showing all results achieved, while Appendix E contains the expert review surveys and their results.

Background & Literature Overview

This chapter introduces the foundational concepts relevant to the study. It surveys the key literature that informs the proposed architecture, drawing from three intersecting areas: Algorithmic Trading, LLMs, and Deep Reinforcement Learning (DRL).

2.1 | Algorithmic Trading

Algorithmic trading has become a dominant theme in modern financial markets, with automated systems responsible for the majority of trade executions across most exchanges. In the United States, algorithmic trading accounts for approximately 60% of total equities trading volume, with High-Frequency Trading (HFT) contributing to nearly half of this activity (Chlistalla, 2011). The growth and prevalence of algorithmic trading can be attributed to three key technological and structural advancements: (1) increased computational power, enabling faster and more efficient execution of complex trading strategies; (2) breakthroughs in financial engineering and machine learning (ML), allowing for more sophisticated predictive models and risk management techniques; and (3) the evolving nature of financial markets, characterized by greater speed, scale, and complexity, which has necessitated the adoption of automated trading strategies to remain competitive (Bartram et al., 2020).

Algorithmic strategies vary in scope, from long-term investment frameworks to extreme-short-term trading models designed to exploit transient market inefficiencies.

2.1.1 | Long-term Investment

Long-term algorithmic trading strategies focus on portfolio optimization, risk-adjusted asset allocation, and systematic investment strategies over timeframes of months or years. Institutional investors utilize modern portfolio theory (MPT), factor-based investing with Factor Models, and other econometric techniques. Factor models, such as the Fama-French Five-Factor Model, attribute an equity's returns to predefined risk

factors. These evolved from the capital asset pricing model (CAPM), which correlates portfolio returns with market risk and was later expanded by Eugene Fama and Kenneth French with their seminal Three-Factor (Fama & French, 2004) and subsequent Five-Factor Model (Fama & French, 2015). These models link companies' characteristics to excess returns through factor weights, which are then used to create portfolios with the goal of delivering above-market returns. The model is estimated by regressing historical portfolio returns against the factor returns, decomposing risk and return components. The Five-Factor Model is formalized as follows (Fama & French, 2015):

$$R_i - R_f = \alpha_i + b_i(R_m - R_f) + s_i\text{SMB} + h_i\text{HML} + r_i\text{RMW} + c_i\text{CMA} + \varepsilon_i \quad (2.1)$$

where R_i is the return of stock i , R_f is the risk-free rate, and R_m is the market return. The model includes multiple risk factors: *SMB* (*Small Minus Big*), which captures the size premium for small stocks over large stocks; *HML* (*High Minus Low*), which measures the excess returns of high book-to-market versus low book-to-market stocks; *RMW* (*Robust Minus Weak*), which reflects the profitability factor by comparing firms with robust versus weak profitability; and *CMA* (*Conservative Minus Aggressive*), which accounts for the tendency of firms with conservative investment policies to outperform those with aggressive investments. The parameter α_i represents the model intercept, and $b_i, s_i, h_i, r_i,$ and c_i are the factor loadings that quantify a stock's sensitivity to each factor, with ε_i being the error term.

The rise of financial machine learning has augmented such traditional econometric models (López de Prado, 2020a). ML introduces nonlinear and high-dimensional statistical methods that often surpass traditional econometric approaches in financial modeling, particularly in advanced applications such as language models (Ding et al., 2024; Liu, Yang, et al., 2022; Mai, 2024; Takara et al., 2023; Théate & Ernst, 2021). However, challenges such as overfitting and false signals remain prevalent, or even intensify, due to the inherently noisy, sparse, and frequently non-stationary nature of financial data (López de Prado, 2018b). Additionally, selection bias and data leakage persist as critical concerns, further complicating model reliability (Joubert et al., 2024).

2.1.2 | Short-term Trading

Short-term algorithmic trading strategies exploit temporary price inefficiencies using statistical arbitrage, mean-reversion, and momentum-based approaches and their time frame is from seconds to days. Statistical arbitrage leverages cointegration techniques, and clustering to identify asset pairs exhibiting mean-reverting behavior.

Mean-reversion trading relies on mathematical and stochastic models, which describes the tendency of a price series to revert to its mean over time (Chan, 2013) as visualized in the figure 2.1.

Momentum-based strategies aim to exploit persistent trends in the time series of asset returns. This approach assumes that an asset's own past return over a fixed horizon positively predicts its near-term future return, a pattern documented across equities, bonds, currencies, and commodities.

A model used to detect this behavior is a sign-based regression (Moskowitz et al., 2012), the returns are scaled by their ex-ante volatility:

$$\frac{r_t^s}{\hat{\sigma}_{t-1}^s} = \alpha + \beta_h \cdot \text{sign} \left(\frac{r_{t-h}^s}{\hat{\sigma}_{t-h-1}^s} \right) + \epsilon_t^s \quad (2.2)$$

Here, r_t^s denotes the excess return of asset s at time t , $\hat{\sigma}_{t-1}^s$ is the ex-ante volatility estimate at time $t - 1$, and h is the lag length set to 12 months. The predictor is the sign of the standardized return lagged h periods. A statistically significant positive coefficient $\beta_h > 0$ indicates that assets with positive past returns tend to continue performing well, consistent with the time series momentum hypothesis.

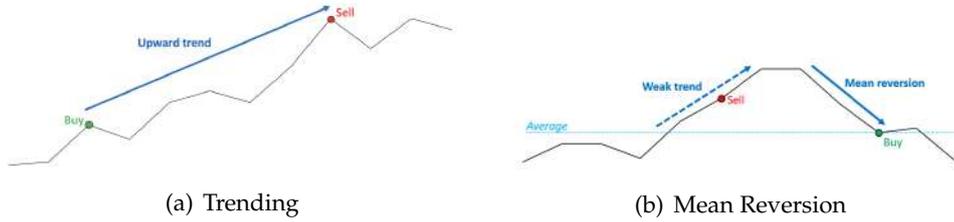


Figure 2.1: Mean Reversion & Trending Visualization (Théate & Ernst, 2021).

If a security exhibits mean-reverting behavior, its reversion speed can be estimated using the discrete Ornstein-Uhlenbeck (OU) process (Chan, 2013), given by:

$$\Delta y(t) = \lambda(y(t-1) - \mu) + \epsilon_t \quad (2.3)$$

where ϵ_t is Gaussian noise, μ represents the long-term mean of the process, and λ determines the strength of mean reversion. Estimating λ via linear regression of $\Delta y(t)$ against $y(t-1)$ allows us to compute the half-life of mean reversion:

$$t_{1/2} = \frac{\ln(2)}{|\lambda|} \quad (2.4)$$

The most extreme form of short-term trading is HFT where proprietary trading firms execute thousands of trades per second, leveraging order book imbalances and arbitrage opportunities across exchanges (Y. Wang & Yan, 2021). Although it improves liquidity in the market, HFT introduces systemic risks. Events such as the 2010 "Flash Crash" demonstrated how rapid algorithmic trading can exacerbate market instability (Chlistalla, 2011). Sánchez Serrano (2020) examines the systemic vulnerabilities associated with high-frequency trading (HFT) and evaluates interventions, such as speed limits, circuit breakers, and restrictions on aggressive trading strategies, aimed at curbing market distortions.

HFT influences market dynamics by reacting to new market data faster than traditional market participants. This speed advantage allows HFT firms to adjust asset prices based on incoming signals, leading to rapid price changes. A simplified model describing this process assumes that the price of a stock P^z at time $t + 1$ is a function of its past value, new information I_t^z , and other residual factors:

$$P_{t+1}^z = P_t^z + \alpha \Delta I_t^z + u_t^z \quad (2.5)$$

where P_t^z is the observed price at time t , ΔI_t^z represents changes in available market information, α captures the sensitivity of price to new information, and u_t^z accounts for other stochastic influences. In calm market conditions, the residual component u_t^z remains low, and price adjustments are minimal. However, during periods of uncertainty or market stress, non-HFT participants may struggle to process information at the same speed, leading to price distortions or liquidity withdrawal. This highlights the dual role of HFT: enhancing price efficiency under normal conditions while potentially exacerbating volatility in stressed environments (Sánchez Serrano, 2020).

Sentiment analysis has become an essential component of algorithmic trading, enabling trading strategies to incorporate market sentiment from financial news, corporate reports, and social media sources. LLMs have significantly advanced sentiment-driven trading by improving the extraction and classification of financial sentiment, allowing for real-time market adaptation (Ding et al., 2024). In News-driven trading LLMs analyze stock-related news and macroeconomic reports to predict short-term price movements (Ding et al., 2024).

In algorithmic trading, evaluating raw returns alone is insufficient to assess a strategy's effectiveness, as high returns may be accompanied by significant risk or volatility. To compare trading algorithms fairly, especially under varying market conditions, performance must be measured in a way that accounts for both return and risk. The Sharpe ratio is a widely adopted risk-adjusted metric that quantifies the excess return per unit

of risk, where risk is typically measured by the standard deviation of returns (Sharpe, 1966), and is given by:

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_p} \quad (2.6)$$

where R_p is the strategy's return, R_f is the risk-free rate, and σ_p is the standard deviation of strategy's returns.

Trust issues also arise from opaque algorithms, black-box AI decisions, and the lack of regulatory clarity, leading to investor hesitancy in deploying capital due to fears of hidden biases or false strategies (Bartram et al., 2020). These systemic trust deficits demand a shift towards models that enhance fidelity and robustness in financial AI applications (Bartram et al., 2020). Unlike traditional AI-driven trading systems, which often prioritize speed and complexity over interpretability, LLM-enhanced frameworks can exhibit human-like rationality, learning from market shifts while maintaining transparency and accountability (Yu et al., 2024). This ensures that algorithmic trading does not merely optimize for profit but adheres to fiduciary principles, fostering investor and trader confidence.

Dakalbab et al. (2024) surveys the application of ML techniques in algorithmic trading from 2015 to 2023, as shown in Figure 2.2, highlighting their distinct advantages in forecasting and trade execution. Among these approaches, RL and DRL have emerged as powerful tools for optimizing trading strategies. Deep learning, found in 30% of studies, enhances market predictions by processing vast amounts of unstructured data to identify complex trading patterns, price movements, and sentiment trends. Unfortunately these deep learning models lack decision-making autonomy and require additional mechanisms for strategy adaptation. Reinforcement Learning (RL), used in 29% of studies, directly addresses this limitation by enabling trading agents to dynamically adjust strategies based on market conditions. Unlike traditional models, RL does not rely on predefined rules; instead, it continuously learns through trial and error, optimizing decisions to maximize long-term rewards. When combined with deep learning, a DRL-powered trading system can dynamically adjust strategies based on market conditions, demonstrating potential in automated trading. While such models show adaptability, their real-time application in time-critical trading systems remains largely unexplored due to hardware constraints and market latency (Pricope, 2021).

Investment firms generally operate under two primary structures: top-down and bottom-up, which define how investment decisions flow. The structure of investment firms influences how algorithmic trading strategies are designed and implemented, particularly in the context of ML and RL applications. Large top-down firms, such as Black-

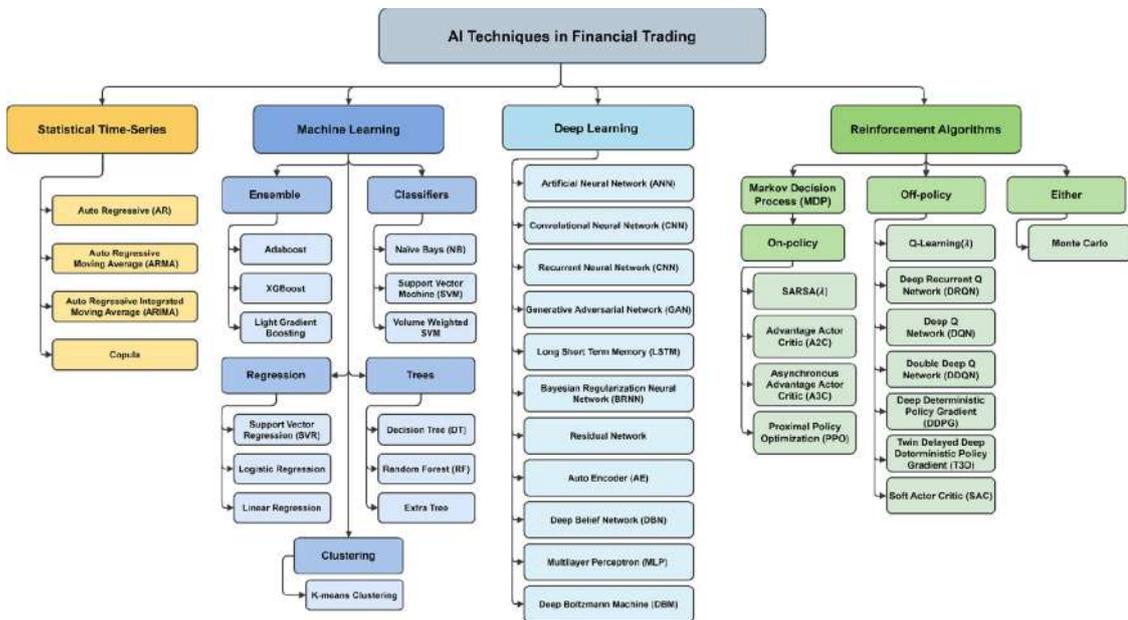


Figure 2.2: AI Techniques utilized Trading (Dakalbab et al., 2024)

Rock or Bridgewater Associates, operate with a centralized hierarchy where the CIO and senior strategists set macroeconomic-driven investment themes that guide Portfolio Managers (PMs). These PMs allocate capital across asset classes while relying on quant researchers and analysts to refine predictive models and extract actionable signals. Execution traders and trading desks then ensure efficient order execution while adhering to risk management constraints. By contrast, bottom-up hedge funds, such as Millennium and Citadel, operate under a decentralized, PM-driven structure where each PM is responsible for their own P&L and relies on independent quant teams and traders to execute strategies tailored to specific market inefficiencies. In these firms, decision-making is localized, allowing for more adaptive, data-driven approaches to trading.

The effectiveness of such trading strategies is ultimately tied to the broader debate on market efficiency and price discovery. A fundamental principle in financial economics, the Efficient Markets Hypothesis (EMH) assumes that asset prices reflect all available information to the market, making it impossible to systematically achieve excess returns (Fama, 1970). The EMH is typically classified into three forms:

- **Weak form:** Prices reflect all historical data, making technical analysis ineffective.
- **Semi-strong form:** Prices incorporate all publicly available information, limiting the effectiveness of fundamental analysis.

- **Strong form:** Prices reflect all private and public information, implying no investor can consistently gain an advantage.

A key implication of the EMH is that asset prices follow a stochastic process, often modeled as a martingale, where future price movements are independent of past trends. In an efficient market, profitable trading strategies based solely on historical or publicly available data should not consistently outperform.

However, empirical research challenges the strict assumptions of the EMH, highlighting inefficiencies caused by cognitive biases, sentiment-driven trading, and institutional constraints (M. Wang et al., 2024). While the EMH remains foundational, the rise of algorithmic trading suggests persistent inefficiencies that advanced models can exploit. This research investigates how RL and LLM-based systems can systematically identify and act on these inefficiencies, potentially challenging traditional market efficiency assumptions.

2.1.3 | Investor Risk Profiles

An investor risk profile represents a structured assessment of an individual's or institution's capacity and willingness to assume financial risk. For individuals, this profile influences investment behavior by guiding asset allocation according to their tolerance for volatility, financial objectives, and time horizon. For institutions or lenders, it informs credit decisions based on a borrower's capacity to manage debt.

Individual risk profiles are commonly categorized along a continuum from conservative to aggressive. Conservative investors tend to prioritize capital preservation and prefer instruments with stable returns, such as government bonds, blue-chip equities, and money market funds. Aggressive investors, by contrast, are more willing to allocate capital to higher-volatility assets with greater return potential, including speculative equities, emerging market exposures, and alternative investments.

In the context of algorithmic trading, investor risk profiles serve as essential constraints on strategy selection. Systems that dynamically adapt to such profiles, whether defined by personality, capital structure, or market sensitivity, can offer more aligned and personalized financial decision-making (Tang et al., 2022).

2.2 | Reinforcement Learning in Trading

Traditionally, ML in finance has focused on supervised learning tasks such as return forecasting, sentiment analysis, and risk classification. Yet these models fail to capture

the temporal dependencies and interactions that characterize real-world trading (Dang, 2020). They optimize one-step predictions without considering the long-term outcomes of sequential decisions.

This limitation has led to the increasing adoption of RL, a learning paradigm designed to handle sequential decision-making under a changing environment. In RL, an agent learns a policy $\pi(a | s)$ to interact with an environment with the goal of maximizing cumulative reward over time. This interaction is formalized as a Markov Decision Process (MDP), defined as $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, P, r, \gamma \rangle$, where \mathcal{S} and \mathcal{A} are finite state and action spaces, $P(s' | s, a)$ defines the transition dynamics, $r(s, a)$ is the reward function, and $\gamma \in [0, 1)$ is the discount factor. The goal in RL is to find a policy $\pi : \mathcal{S} \rightarrow \Delta(\mathcal{A})$ that maximizes the expected return (Sutton & Barto, 2018):

$$J(\pi^*) = \max_{\pi} \mathbb{E}_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t) \right] \quad (2.7)$$

subject to $s_{t+1} \sim P(\cdot | s_t, a_t)$ and $a_t \sim \pi(\cdot | s_t)$. Q-learning is a foundational algorithm in RL, representing a model-free, off-policy learning method that estimates the optimal action-value function $Q^*(s, a)$ without requiring prior knowledge of the environment's transition dynamics (Sutton & Barto, 2018). The core principle of Q-learning is derived from the recursive structure of the Bellman optimality equations. These equations characterize the value of a state or state-action pair as the expected sum of immediate rewards and discounted future returns, optimized over all possible actions:

$$V^*(s) = \max_a \sum_{s', r} P(s', r | s, a) [r + \gamma V^*(s')], \quad (2.8)$$

$$Q^*(s, a) = \sum_{s', r} P(s', r | s, a) \left[r + \gamma \max_{a'} Q^*(s', a') \right]. \quad (2.9)$$

The optimal policy π^* can then be obtained by acting greedily with respect to $Q^*(s, a)$, selecting actions that maximize expected returns at each state. A key strength of Q-learning lies in its off-policy nature: it can converge to $Q^*(s, a)$ even when learning from trajectories generated by an exploratory (non-optimal) behavior policy (Sutton & Barto, 2018).

In practice, the agent updates its estimate of the action-value function using temporal-difference learning, according to the rule:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r_t + \gamma \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t) \right], \quad (2.10)$$

where $\alpha \in (0, 1]$ denotes the learning rate, $\gamma \in [0, 1)$ is the discount factor, and r_t is the reward observed after executing action a_t in state s_t . The term $\max_{a'} Q(s_{t+1}, a')$ reflects the agent's estimate of the highest attainable return from the next state, assuming greedy action selection going forward.

The convergence of Q-learning to the optimal value function has been established under standard assumptions in the tabular setting. Moreover, its conceptual framework underpins many modern DRL algorithms (Mnih et al., 2015).

To illustrate these ideas in a domain-specific context, consider a long-short trading environment involving a single financial instrument. In this setting, an RL agent selects either a long or a short position at each decision point, with the goal of maximizing cumulative returns over time. The environment transitions between discrete market regimes, such as upward, downward, or sideways trends-based on recent price dynamics. These dynamics can be modeled as a Markov Decision Process (MDP), where the next regime depends only on the current regime and the agent's action, satisfying the Markov property. This simple financial scenario provides an accessible, finite-state instantiation of tabular Q-learning and exemplifies its applicability to sequential decision-making under uncertainty.

A limitation of tabular Q-learning lies in its inability to scale to high-dimensional or continuous environments (Sutton & Barto, 2018). As the number of state-action pairs grows, storing and updating individual Q-values becomes computationally infeasible. To address this, function approximation techniques are employed to generalize across similar states and actions. A known solution is the Deep Q-Network (DQN), which integrates deep learning with reinforcement learning by approximating the Q-function using a neural network (Mnih et al., 2013, 2015). Rather than maintaining an explicit lookup table, the agent learns a mapping from state representations to Q-values, enabling efficient learning in complex environments without relying on tabular representations. The DQN is part of a broader class of DRL algorithms known as critic-only methods. This classification, which also includes actor-only and actor-critic families, has been adopted in the context of financial applications by Pricope (2021). These categories are defined by the presence or absence of an explicit policy module and the mechanism by which the agent evaluates and selects actions.

As mentioned, the critic is typically implemented as a parameterized function approximator using a neural network. Mnih et al. (2013, 2015) critic is a convolutional neural network that predicts Q-values from high-dimensional inputs, such as raw image frames. To stabilize training, DQN incorporates two key mechanisms: experience replay and a target network. The target network, parameterized by θ^- , is a delayed copy of the online network θ , used to compute the TD target as

$$y_t = r_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta^-), \quad (2.11)$$

which prevents the target from shifting too rapidly during learning. Experience replay, on the other hand, stores past transitions (s_t, a_t, r_t, s_{t+1}) in a buffer and samples mini-batches uniformly to break correlations between sequential data and improve sample efficiency (Mnih et al., 2015). The target DQN algorithm is described in Appendix 1.

The networks' learning process is based on temporal-difference (TD) learning, a class of methods that update value function estimates based on bootstrapped predictions. In the tabular setting, TD learning updates the value estimate for a state-action pair by taking a step toward the Bellman target. This core principle extends to the function approximation setting used in deep reinforcement learning, where the action-value function $Q(s, a; \theta)$ is parameterized by a neural network. The Deep Q-Network is trained by minimizing the squared temporal-difference error (Mnih et al., 2015), defined for a single transition (s_t, a_t, r_t, s_{t+1}) as

$$\delta_t = y_t - Q(s_t, a_t; \theta), \quad (2.12)$$

where the target y_t is computed using the Bellman backup with a fixed target network in Equation 2.11. The squared TD error is then used to define the loss function:

$$\mathcal{L}(\text{target}) = \mathbb{E}_{(s,a,r,s') \sim \mathcal{D}} \left[(y_t - Q(s, a; \theta))^2 \right], \quad (2.13)$$

where \mathcal{D} denotes the experience replay buffer. The expectation is approximated by sampling a minibatch of past transitions, enabling the network to be updated using stochastic gradient descent (Mnih et al., 2015).

The DQN architecture has a specific limitation, as identified by van Hasselt et al. (2015), whereby the use of the maximization operator in the TD target equation 2.11, can lead to overestimation bias. This arises because the same network is used both to select the next action and to evaluate its value. To mitigate this, Double Deep Q-Network (DDQN) decouples the two roles: the action is selected using the online network, while its value is evaluated using the target network. The modified target becomes

$$y_t^{\text{DDQN}} = r_t + \gamma Q \left(s_{t+1}, \arg \max_{a'} Q(s_{t+1}, a'; \theta); \theta^- \right), \quad (2.14)$$

where θ and θ^- are the parameters of the online and target networks, respectively. This formulation reduces overoptimistic value estimates and leads to more stable and accurate learning, particularly in stochastic environments (Lazaridis et al., 2020).

The DDQN algorithm is described in Appendix 2.

A further refinement to critic-only architectures is the dueling network architecture, which decouples the estimation of the state-value and advantage functions. Instead of learning a single stream to approximate the action-value function $Q(s, a)$, the network is split into two separate estimators: one for the state-value function $V(s)$, and another for the advantage function $A(s, a)$. These are combined through a differentiable aggregation module. This decomposition is particularly effective in states where the impact of the chosen action is minimal, allowing the critic to generalize more effectively and improving training stability (Lazaridis et al., 2020).

An additional enhancement to critic modeling is introduced through distributional reinforcement learning. In contrast to conventional approaches that approximate only the expected return, distributional methods learn a full probability distribution over possible returns for each state-action pair. This richer signal enables more nuanced value estimates and has been shown to improve empirical performance. The distribution is modeled over a fixed set of discrete support values, with a projection step used to maintain the categorical structure during training (Lazaridis et al., 2020).

Actor-only methods constitute a distinct class of deep reinforcement learning algorithms, in which the agent learns a parameterized policy directly without estimating a value function. These methods optimize the expected return by performing gradient ascent on the parameters of a stochastic policy $\pi_\theta(a | s)$. The foundational algorithm in this category is REINFORCE, which applies the policy gradient theorem to update parameters in the direction of increased likelihood of high-return actions. REINFORCE computes gradients using sampled returns from full episodes, resulting in unbiased but high-variance updates (Lazaridis et al., 2020).

To mitigate this variance, a baseline function, typically the state-value function $V^\pi(s)$, can be subtracted from the return, yielding an advantage estimate that stabilizes learning without introducing bias. Still, actor-only methods remain sensitive to learning rates, reward scaling, and trajectory length, which limits their practicality in environments with sparse rewards or delayed feedback. Lazaridis et al. (2020) states that actor-only approaches are conceptually simple and well-suited to continuous action spaces, but they are less prevalent in modern practice due to their inefficient use of data and the lack of stabilizing feedback from a learned value function.

To address these limitations, second-order optimization methods have been explored. These approaches aim to improve the stability and convergence of policy gradient updates by incorporating information about the curvature of the policy space. One such method is Trust Region Policy Optimization (TRPO), which introduces a constraint on the step size in policy space using a Kullback–Leibler (KL) divergence penalty. Rather than allowing large, potentially destabilizing updates, TRPO restricts the policy change

to remain within a trust region. While this improves stability, Lazaridis et al. (2020) note that TRPO is computationally expensive due to the need to estimate second-order derivatives and solve a constrained optimization problem at each step.

Actor-only architectures are best applied in environments where rewards are frequent and immediate, and where the overhead of training a critic is unjustified. In practice, though, they are often outperformed by hybrid methods that combine the flexibility of policy gradients with the evaluative feedback of value functions (Lazaridis et al., 2020).

Actor-critic methods combine the advantages of actor-only and critic-only architectures by jointly learning a parameterized policy and a value function. The actor, typically a stochastic policy $\pi_{\theta}(a | s)$, is responsible for selecting actions, while the critic estimates expected returns to guide the policy update. The critic may be implemented as either a state-value function $V(s)$ or an action-value function $Q(s, a)$, with both components parameterized and trained concurrently. This architecture allows for more stable and efficient learning by reducing the variance associated with policy gradient estimates through bootstrapping from the critic.

Actor-critic methods are widely adopted due to their balance between exploration and exploitation, as well as their ability to operate effectively in both discrete and continuous action spaces (Lazaridis et al., 2020). The critic provides evaluative feedback for the actor, enabling sample-efficient updates, while the actor introduces the flexibility needed to learn complex policies. This architecture underpins many state-of-the-art algorithms.

Among the early scalable actor-critic algorithms are Advantage Actor Critic (A2C) and Asynchronous Advantage Actor Critic (A3C). Both methods estimate policy gradients using n -step returns and employ advantage estimates to reduce variance. A2C performs synchronized updates using batched rollouts across multiple environments, while A3C enables asynchronous training across parallel workers, facilitating faster convergence and improved exploration.

A more recent and widely adopted variant is Proximal Policy Optimization (PPO), which builds on the principles of policy gradient optimization while introducing a clipping mechanism to constrain policy updates. Rather than relying on second-order derivatives or hard trust-region constraints, PPO restricts the deviation between new and old policies via a clipped surrogate objective. This regularization improves stability without incurring the computational overhead of constrained optimization. PPO has become one of the most widely used reinforcement learning algorithms due to its robustness, ease of implementation, and strong empirical performance across a broad range of tasks (Lazaridis et al., 2020; Li, 2017; Liu, Yang, et al., 2022).

Off-policy actor-critic methods extend the basic actor-critic framework by decoupling the data collection policy from the learning policy, enabling improved sample efficiency and more flexible reuse of past experiences. These methods are particularly well suited to continuous action spaces and tasks with high sample complexity, where on-policy approaches such as PPO may be inefficient or unstable. A prominent example is Deep Deterministic Policy Gradient (DDPG), which learns a deterministic policy function and employs experience replay and target networks to stabilize training (Lazaridis et al., 2020). The actor is updated via gradients propagated through the critic, which approximates the action-value function. To ensure sufficient exploration, noise is added to the policy during training.

While DDPG represented a significant advance in continuous control, it is sensitive to overestimation bias and function approximation error. To address these limitations, Twin Delayed DDPG (TD3) was introduced as an extension. TD3 incorporates multiple improvements: the use of two critics and taking the minimum of their predictions to reduce overestimation, delaying actor updates to allow for more accurate critic estimates, and adding noise to the target policy during bootstrapping. These modifications result in a more stable and robust learning process.

Soft Actor-Critic (SAC) further develops the off-policy actor-critic paradigm by incorporating entropy maximization directly into the objective function. SAC maintains a stochastic policy and learns not only to maximize expected returns but also to favor policies with higher entropy (Lazaridis et al., 2020). This encourages persistent exploration and leads to improved performance, particularly in environments with sparse or delayed rewards. SAC also builds on TD3's architectural insights, employing two critics and target smoothing to enhance learning stability.

As highlighted by Lazaridis et al. (2020), Q-learning and its deep variants have gained widespread adoption due to their structured approach and empirical performance. Their discrete action-space formulation aligns well with standardized evaluation environments such as the Atari Learning Environment (ALE), supporting reproducibility and consistent benchmarking. Furthermore, these methods allow direct estimation of $Q(s, a)$, facilitating structured learning and value assessment. Empirically, DDQN significantly improves upon the original DQN by mitigating overestimation bias through the decoupling of action selection and evaluation. DDQN also serves as a foundation for more advanced value-based architectures such as Dueling DQN, C51, and Rainbow, reflecting the cumulative benefits of successive refinements.

However, as Lazaridis et al. (2020) notes, Q-learning methods, including DDQN, have limitations in continuous control tasks where actor-critic approaches like Proximal Policy Optimization (PPO), TD3, and Soft Actor Critic (SAC) demonstrate superior

adaptability and sample efficiency. These algorithms benefit from flexible policy representations and stabilization techniques such as entropy regularization, delayed updates, and trust-region constraints, which make them more effective in high-dimensional action spaces.

While model-based methods offer potential improvements in sample efficiency, they have yet to outperform model-free alternatives in complex benchmarks such as ALE. The difficulty of learning accurate long-horizon predictions from pixel-based observations and the accumulation of model errors continue to hinder their effectiveness (Lazaridis et al., 2020). As a result, DDQN remains an important baseline, balancing architectural simplicity, reproducibility, and strong empirical performance.

2.3 | LLMs in Decision-Making

LLM represent a class of models capable of learning rich statistical representations over text corpora at scale. Their foundational architecture is the transformer, introduced by Vaswani et al. (2017), which drops recurrence in favor of self-attention mechanisms. An upgrade over Recurrent Neural Network (RNN), transformers process all tokens simultaneously, allowing each token to attend to every other token in the input sequence. This design enables efficient parallel computation and overcomes the vanishing gradient issues that plagued prior architectures.

The application of LLM to financial decision-making has followed multiple paths. One path involves generative models trained directly on price data. Mai (2024) introduces StockGPT, an autoregressive transformer model trained on over 70 million daily U.S. stock returns. By discretizing continuous returns into tokens and modeling them as sequences, the model predicts future returns autoregressively, capturing latent market dynamics.

Another direction involves LLM-based sentiment analysis from financial text. Zhou et al. (2021) show that BERT and RoBERTa models trained on headline sentiment outperform traditional methods in predicting short-term stock movements. Similarly, Lopez-Lira and Tang (2023) demonstrate that sentiment signals generated by ChatGPT can be used to construct statistically significant trading strategies.

In the agent-based paradigm, Ding et al. (2024) identify architectures such as FinAgent, FinMem, and AlphaGPT, which use LLM as central components for decision-making, memory reflection, and alpha factor generation. These systems incorporate not only textual inputs but also numerical time-series and structured market data. GPT-4, Claude, and LLaMA 2 are all prominently used across these agents.

Empirical results in Mai (2024) show that when trained on long sequences of historical returns, StockGPT can recover well-known patterns such as momentum and reversal effects. Its ability to forecast returns persists over both monthly and daily horizons, suggesting that the attention mechanism effectively models temporal dependencies and regime shifts. Moreover, financial-specific LLMs such as FinGPT and BloombergGPT encode long-range textual relationships across documents and events. These models benefit from transformers' capacity to link economic narratives spanning annual reports, earnings guidance, and macroeconomic releases (Ding et al., 2024). According to Cao et al. (2024), LLMs can be enhanced with external tools, long-term memory, and role conditioning to support domain-specific reasoning. FinAgent (Zhang et al., 2024) and FinMem (Yu et al., 2024), for instance, employ layered memory to persist and retrieve relevant information across market episodes, enabling coherent responses to sequential trading tasks.

From the latest empirical research on LLM, transformer-based models and fine-tuned financial variants like, have demonstrated substantial effectiveness in financial sentiment analysis. Their utility is particularly evident in long-short equity strategies, where sentiment scores derived from financial news or social media are employed to rank assets and construct portfolios. This methodology is supported by studies such as Lopez-Lira and Tang (2023) and Mai (2024), which highlight the predictive capacity of LLM-based sentiment signals in real-time trading scenarios. Notably, between 2023 and 2024, comparative evaluations have consistently positioned Generative Pretrained Transformer (GPT)4 as the most performant model across a spectrum of financial tasks, from sentiment forecasting to trading signal generation. As illustrated in Table 2.1, GPT4 outperforms both closed- and open-source alternatives, reaffirming its leading role in financial applications of generative AI.

It is worth noting that general-purpose benchmarking studies also identify GPT-4 as the most performant model across a wide range of natural language processing benchmarks. In particular, Naveed et al. (2024) provide a systematic overview of leading LLM, comparing architectures such as PaLM, LLaMA2, OPT, Claude, Gemini, and DeepSeek. Their findings consistently place GPT-4 at the top in terms of reasoning, multilinguality, and instruction following.

For LLM to perform effectively, it is critical to construct prompts that align with the model's learned representations and task requirements. Prompt engineering refers to the systematic design and optimization of input sequences called prompts that condition the behavior of LLMs for specific tasks. This process leverages the transformer architecture's reliance on self-attention mechanisms, which encode dependencies across the entire prompt and output sequence (Vaswani et al., 2017). Consequently, the struc-

Paper	Models Tested	Best Model
<i>Can ChatGPT Forecast Stock Price Movements?</i> (Lopez-Lira & Tang, 2023)	GPT-1, GPT-2, GPT-3.5, GPT-4, BERT, FinBERT	GPT-4
<i>FINMEM: A Performance-Enhanced LLM Trading Agent</i> (Yu et al., 2024)	GPT-3.5-Turbo, GPT-4, GPT-4-Turbo, davinci-003, Llama2-70b-chat	GPT-4
<i>LLMFactor: Extracting Profitable Factors</i> (M. Wang et al., 2024)	GPT-3, GPT-3.5, GPT-4	GPT-4
<i>StockGPT: A GenAI Model for Stock Prediction and Trading</i> (Mai, 2024)	GPT-4, FinBERT	GPT-4
<i>Evaluating LLMs in Financial Tasks - Code Generation in Trading Strategies</i> (Alonso & Dupouy, 2024)	GPT-4-Turbo, Codellama-70B, Gemini-Pro, Mistral-7B	GPT-4-Turbo
<i>Designing Heterogeneous LLM Agents for Financial Sentiment Analysis</i> (Xing, 2025)	GPT-3.5, GPT-4, BERT, RoBERTa, Mistral	GPT-4
<i>Large Language Model Agent in Financial Trading: A Survey</i> (Ding et al., 2024)	FinGPT, OPT, GPT-4, GPT-3.5, LLaMA2, Baichuan, Qwen	GPT-4

Table 2.1: Evaluation of LLM Performance in Financial Research (2022–2025).

ture and content of a prompt significantly influence the model’s output distribution and performance.

Schulhoff et al. (2024) provides a formal understanding of prompting techniques. Let T be a prompt and Q a question, then the output distribution over an answer $A = (a_1, \dots, a_{|A|})$ is defined as:

$$p(A | T, Q) = \prod_{i=1}^{|A|} p_{\text{LM}}(a_i | T, Q, a_1, \dots, a_{i-1}),$$

where p_{LM} is the pretrained language model. This formulation captures the autoregressive nature of most LLMs, where each token is predicted conditioned on both the prompt and prior tokens.

In practical applications, prompts are typically instantiated through a prompt template, denoted $T(x^*)$, which maps a data example $x^* \in D_{\text{eval}}$ to a concrete prompt. In this context, the prompt is considered as an instantiation of a template function $T(\cdot)$, which takes a data point as input and returns the corresponding prompt. Reformulating the conditional distribution using a prompt template yields:

$$p(A | T(x^*)) = \prod_{i=1}^{|A|} p_{\text{LM}}(a_i | T(x^*), a_1, \dots, a_{i-1}),$$

where $x^* \in D_{\text{eval}}$ is an item from the evaluation dataset. This reformulation emphasizes the use of a generalizable function $T(\cdot)$, rather than a fixed string, in constructing prompts. In many settings, particularly in few-shot prompting, the prompt includes not only the target instance x^* but also a set of exemplars $D_{\text{train}} = \{(x_1, y_1), \dots, (x_n, y_n)\}$ drawn from a similar distribution:

$$X = \{x_1^*, x_2^*, \dots, x_m^*\}.$$

These exemplars are inserted into the prompt to guide the model via In-Context Learning (ICL), a key capability by which LLMs perform meta-learning without weight updates. This phenomenon is understood to arise from the model implicitly performing gradient-like computations in its activation space (Dai et al., 2023b).

A systematic taxonomy of prompting techniques, as developed by Schulhoff et al. (2024), identifies 58 text-based methods grouped into six high-level categories: In-Context Learning (ICL), Thought Generation, Decomposition, Ensembling, Self-Criticism, and Zero-Shot Prompting. These categories encompass both simple formats like direct instructions and sophisticated strategies involving iterative refinement, logical chaining, or ensemble aggregation.

Despite their potential, LLMs face significant challenges in financial forecasting and decision support. Ding et al. (2024) note that LLM-based trading agents often hallucinate or misinterpret financial content, particularly when generating trading signals or decision support outputs. These issues can be partially mitigated through memory-augmented architectures, such as FinMem, which enhance context retention and reasoning capabilities.

Evaluating these systems remains difficult, as financial predictions involve high noise and lack a clear ground truth. Unlike classification tasks with well-defined accuracy metrics, financial markets exhibit non-stationary dynamics, making performance assessment inherently complex (Ding et al., 2024).

The sensitivity of LLMs to prompt formulation introduces substantial variability in their outputs, which undermines the stability required in decision-critical systems. The meta-analysis by Schulhoff et al. (2024) demonstrates that even minor syntactic variations in prompts can lead to divergent model behaviors-particularly under few-shot and chain-of-thought prompting regimes. In financial domains, where consistent and reproducible reasoning is essential for compliance, auditability, and institutional trust,

such fragility poses a significant operational risk. Addressing this challenge requires structured methods of prompt mining and optimization (Gonen et al., 2023; Schulhoff et al., 2024; S. Wang et al., 2024).

2.4 | Long-Term Planning Challenges in RL

Effective long-term planning remains a challenge in RL, and is known as the credit assignment problem (Sutton & Barto, 2018). Most value-based methods, including DQN and its variants, rely on short-horizon temporal-difference (TD) updates that are structurally constrained in their ability to propagate delayed consequences or capture temporally abstract patterns (Arulkumaran et al., 2017; Pateria et al., 2021).

Arulkumaran et al. (2017) identify long-horizon reasoning as a core unsolved problem in deep RL, highlighting the instability of bootstrapped value propagation, the tendency to overfit to immediate reward signals, and the erosion of distinctions between distant states. These challenges are exacerbated in financial domains, where rewards are often sparse, delayed, and embedded in stochastic and non-Markovian dynamics (Takara et al., 2023). Takara et al. (2023) show that DQN-based agents tend to optimize short-term reward proxies, but fail to generalize across regime shifts or volatility events. Reward design itself further contributes to brittleness, as Booth et al. (2023) report that misalignment between modeled returns and higher-level strategy objectives can lead to policy collapse or overfitting to spurious correlations in the data, particularly when exploration is limited or the training environment lacks sufficient diversity.

Hierarchical Reinforcement Learning (HRL) offers a principled response to these limitations via temporal abstraction. As surveyed by Pateria et al. (2021), HRL decomposes complex tasks into subgoals or callable options, enabling agents to reason at multiple time scales and improving both credit assignment and strategic consistency. In open-ended environments such as Minecraft, Yuan et al. (2023) demonstrate that integrating low-level skill learning with high-level planning substantially improves both sample efficiency and task completion rates.

Deploying an LLM is a newer but promising alternative to manually engineered hierarchies. Instead of acquiring temporal abstractions through exploration alone, LLMs can be prompted to generate symbolic decompositions of long-horizon tasks. For example, Dalal et al. (2024) introduce a framework in which LLMs produce ordered subgoal sequences, constraining the planning space for downstream RL agents. Similarly, Xie et al. (2023) leverage natural language descriptions to synthesize structured reward functions, thus circumventing unstable TD learning by injecting external semantic priors.

Together, these findings suggest that effective long-term planning in RL often benefits from auxiliary scaffolding—whether hierarchical or symbolic—that compensates for the limitations of end-to-end value learning. Rather than relying exclusively on TD updates, agents achieve more robust performance when guided by high-level abstractions grounded in domain structure (Pateria et al., 2021).

To address the core limitations of deep value-based learners in long-horizon contexts, namely, representational collapse, inadequate subgoal discovery, and temporal myopia, state-of-the-art solutions increasingly rely on hierarchical decomposition or LLM-generated priors. These approaches are particularly effective in environments characterized by delayed rewards, partial observability, or complex causal structure (Pateria et al., 2021).

2.5 | Related Works

2.5.1 | Deep Reinforcement Learning (DRL)

The paper Théate and Ernst (2021), “An Application of Deep Reinforcement Learning to Algorithmic Trading” investigates the application of Q-Learning in algorithmic trading using a DDQN. Historical market data spanning 2012 to 2019 is employed, with detailed experiments on Apple (AAPL) and Tesla (TSLA). The model learns optimal trading strategies through simulated interactions with the market, using returns as rewards to train the agent. The achieved SR are 1.484 for AAPL and 0.261 for TSLA.

An aspect of their RL approach is the control and reduction of the action space and traded amounts, which helps prevent unstable learning dynamics and unrealistic policies. This is achieved by discretizing the action space into predefined increments, limiting the number of shares that can be bought or sold at each step. To ensure feasibility and avoid excessive risk exposure, the traded quantities are upper-bounded by:

$$|Q| \leq \frac{v_c}{p_t(1+C)}$$

where v_c is the available capital, p_t the current asset price, and C the trading cost fraction. Beyond the upper bound on traded quantities, the paper also imposes a lower bound based on the requirement that the agent must always maintain a sufficient reserve to close short positions, constrained by the estimated maximum market movement. Specifically, the cash balance must be large enough to cover the worst-case price fluctuation scenario, leading to an additional constraint:

$$v_c \geq -n_t p_t (1 + \epsilon)(1 + C)$$

where n_t represents the number of shares currently held, and ϵ models the maximum expected relative price variation, of which they include trading costs, spreads and slippage. These constraints ensure that the agent remains solvent and can manage risk even in extreme market conditions.

Another key element discussed is the tuning of the discount factor γ , which determines the weight given to future rewards. In their approach, the authors highlight that a lower discount factor ($\gamma \rightarrow 0$) leads to myopic behavior, where the agent prioritizes immediate gains and ignores long-term market trends. On the other hand, excessively high values ($\gamma \rightarrow 1$) may cause the agent to over-prioritize future gains at the cost of suboptimal short-term decisions, potentially making it slow to react to shifts in the market or the traded security.

The TQN improves upon the standard DQN by mitigating overestimation bias. It does this through double Q-learning and improving sample efficiency via prioritized replay. Their results indicate that these refinements yield more stable policies, particularly in volatile financial markets.

In Lazaridis et al. (2020), the review lists three primary categories for RL: model-free, model-based, and modular approaches. Model-free methods, such as DQN and Policy Gradient methods, learn optimal policies directly from interactions with the environment, with extensions like Dueling DQN and Actor-Critic methods improving stability and sample efficiency. Model-based approaches, including probabilistic inference techniques like PILCO and latent-state-based methods such as PlaNet, leverage learned environment dynamics to enhance decision-making. Modular architectures, exemplified by hierarchical and unsupervised RL, address generalization, transfer learning, and multi-task capabilities, facilitating broader AI applications. Recent advancements such as distributed RL frameworks, including Ape-X and IMPALA, have demonstrated scalability improvements. Moreover, exploration strategies, including curiosity-driven methods and count-based exploration, have enhanced sample efficiency in sparse-reward environments.

To provide an empirical evaluation, the authors compiled results from a wide range of benchmark tests, particularly focusing on the Arcade Learning Environment (ALE) and the MuJoCo physics simulation. These tests demonstrated the varying effectiveness of RL algorithms across different environments, highlighting trade-offs between sample efficiency, stability, and final performance. The comparison figures and tables illustrate that hybrid approaches often outperform pure model-free or model-based methods, emphasizing the necessity of combining structured planning with deep policy learning. The structured comparison of RL methods in Lazaridis et al. (2020) provides a valuable baseline for assessing how LLM integration in this research can perform.

The FinRL platform is presented in Liu, Rui, et al.; Liu, Yang, et al.; Liu et al. (2022, 2022, 2021), “FinRL-Meta: A Universe of Near-Real Market Environments for Data-Driven Deep Reinforcement Learning in Quantitative Finance”; “FinRL: A Deep Reinforcement Learning Library for Automated Stock Trading in Quantitative Finance”; “FinRL: Deep Reinforcement Learning Framework to Automate Trading in Quantitative Finance” and provides structured environments, frameworks, and models for developing and testing reinforcement learning (RL) agents in quantitative finance. These frameworks emphasize the importance of realistic market simulations, standardized backtesting procedures, and modular architectures to facilitate reproducibility and extensibility.

Arulkumaran et al. (2017) provide a comprehensive survey of deep reinforcement learning (DRL), tracing the evolution of RL methods from classical tabular techniques to modern deep function approximators. A central theme of the paper is the fragility of deep value-based methods, particularly those relying on recursive Bellman backups, in environments characterized by partial observability, sparse rewards, and delayed credit assignment. The authors detail how temporal-difference (TD) learning-while sample-efficient-suffers from compounding estimation errors, overestimation bias, and instability when applied with deep neural networks. Their analysis underscores the central tension in DRL between function approximation power and the need for stable, long-horizon credit propagation, particularly in domains where feedback is noisy or delayed.

In articulating these challenges, Arulkumaran et al. (2017) emphasize the inadequacy of one-step updates in temporally extended decision-making problems. They highlight that real-world environments often exhibit strong temporal correlations and long-range dependencies, which break the Markov assumptions required for Q-learning and other value-based methods. Furthermore, they note that standard DQN-style agents fail to model such dependencies due to their reliance on feedforward architectures and scalar TD targets. As a response, the authors review hierarchical and recurrent extensions to deep RL, such as Deep Recurrent Q-Networks (DRQN) and actor-critic methods, which attempt to incorporate temporal abstraction and internal state. Nonetheless, the paper concludes that overcoming the long-term planning gap remains an open problem, particularly in complex, high-stakes environments such as finance, where delayed effects and regime shifts dominate the reward landscape.

Booth et al. (2023) provide a rigorous empirical and user-centric investigation into the systemic failures of reward design in reinforcement learning, demonstrating that trial-and-error reward engineering, a dominant practice among RL practitioners, frequently leads to both overfitting and task misalignment. Their findings show that reward functions can be unintentionally overfit to specific algorithms and hyperparameters, yielding policies that optimize surrogate signals rather than the true task objective.

Through computational experiments and a controlled user study, they show that even expert designers routinely construct reward functions that fail to encode the intended task, especially when adopting myopic, state-based heuristics disconnected from the RL agent’s actual objective of cumulative discounted return. This reward misdesign is compounded by the lack of structural guidance in value-based systems like DQN, where policies are induced from scalar TD error without trajectory-level feedback. By identifying reward function overfitting as a source of brittle generalization and misleading evaluation, the work provides strong evidence for the structural fragility of TD learning and underscores the need for architectures, such as ones with LLM guides, that decouple task specification from low-level reward shaping.

2.5.2 | Large Language Model (LLM)

J. Huang and Chang (2023), “Towards Reasoning in Large Language Models: A Survey” investigates reasoning in large language models (LLMs), analyzing their ability to perform complex cognitive tasks such as problem-solving, decision-making, and logical inference. The paper evaluates multiple reasoning paradigms and techniques, focusing on methods for eliciting and improving reasoning capabilities in LLMs. The study categorizes reasoning in LLMs into several approaches, with Chain of Thought (CoT) prompting identified as a key mechanism for enhancing multi-step reasoning. The survey presents studies demonstrating that prompting LLMs with intermediate reasoning steps significantly improves performance in arithmetic, commonsense, and symbolic reasoning tasks. Simple instructions such as “Let’s think step by step” can activate latent reasoning capabilities without requiring additional training.

Rationale engineering further refines reasoning by curating structured examples that enhance logical inference. The survey highlights complexity-based prompting, where more detailed and hierarchical rationales lead to improved problem-solving performance, as well as automated methods for selecting diverse and representative reasoning exemplars to enhance generalization across tasks.

Problem decomposition is another technique examined, where complex problems are broken into simpler subproblems, improving LLMs’ ability to handle multi-step decision processes. An example of this is least-to-most prompting, which guides models to iteratively solve problems by leveraging previously computed sub-results.

Beyond prompting-based methods, bootstrapping and self-improvement techniques allow LLMs to iteratively refine their reasoning capabilities. The survey describes the Self-Taught Reasoner (STaR) framework, in which models generate their own rationales, filter out incorrect ones, and retrain on high-quality reasoning sequences. It also demon-

strates that LLMs can self-improve without human supervision by enforcing logical consistency across multiple reasoning paths.

The study assesses reasoning-enhanced training approaches, where LLMs are fine-tuned on datasets containing explicit logical structures. The survey highlights findings showing that instruction tuning on diverse reasoning tasks significantly enhances problem-solving abilities and that continual pretraining on structured logical data improves quantitative and mathematical reasoning performance.

Despite these advancements, the study identifies limitations in LLM reasoning capabilities, noting that while LLMs perform well in structured reasoning tasks, they struggle with complex planning problems that humans solve effortlessly. The survey references the FOLIO dataset, which tests first-order logic reasoning in LLMs and reveals significant gaps in logical consistency. It also highlights findings indicating that LLMs often fail at counterfactual reasoning, an essential aspect of robust logical inference.

Yu et al. (2024), “FinMem: A Performance-Enhanced LLM Trading Agent with Layered Memory and Character Design” introduces the FINMEM framework, an LLM-based trading agent that enhances financial decision-making through a structured layered memory system and character design. The framework is designed to overcome limitations in existing LLM-based trading models by integrating multi-source financial data with a hierarchical memory retrieval mechanism, allowing the agent to adapt dynamically to market conditions.

FINMEM’s architecture consists of three core modules: profiling, memory, and decision making. The profiling module defines the agent’s character by embedding domain-specific knowledge and customizable risk inclinations, enabling it to simulate risk-seeking, risk-averse, or self-adaptive trading behaviors. The decision-making module leverages memory-augmented contextual prompts to generate investment actions based on retrieved insights.

The memory module is central to FINMEM’s performance, structuring information into a layered long-term memory system that prioritizes financial data based on time sensitivity and relevance. The memory hierarchy consists of a Shallow memory, which stores recent market news with immediate trading relevance. An intermediate memory, which retains quarterly financial reports for mid-term decision-making. And finally, a Deep memory layer which archives long-term financial statements such as annual reports.

The training process of FINMEM is designed to generate memory references to be stored in their vector database, enabling the LLM to retrieve relevant historical events efficiently. During training, the system processes incoming financial data and classifies it based on its temporal impact. Reflection prompts extract investment insights, which are

stored in memory layers with assigned decay rates to simulate the varying significance of financial information over time. Critical memory events are progressively moved to deeper layers, ensuring that high-impact financial insights remain accessible for extended periods.

The retrieval process in FINMEM is governed by a retrieval-augmented generation (RAG) mechanism, where the agent retrieves the most relevant past events to enrich its decision-making process. Memory retrieval is scored using a combination of recency, relevance, and importance, dynamically adjusting the selection of top-ranked memory events for each trading inquiry. FINMEM refines this retrieval process by incorporating an access counter, which promotes frequently referenced events to deeper memory layers, preserving key insights for long-term strategy formulation.

Experimental evaluations demonstrate that FINMEM achieves superior trading performance with significantly shorter training durations compared to other algorithmic trading agents. Training on just six months to a year of financial data is sufficient to yield robust trading strategies, outperforming deep reinforcement learning (DRL) agents trained on significantly longer historical data. The structured memory layer enables FINMEM to effectively capture, retain, and retrieve high-value financial information, reducing the need for extensive retraining.

Schulhoff et al. (2024) provides a systematic review of prompting techniques in generative language models, introducing a taxonomy that categorizes prompting strategies based on their objectives, structures, and evaluation methodologies.

For text generation, the study delineates prompting methods into families that guide models in optimizing coherence, factuality, and reasoning capabilities. Among these:

- **Instruction-based prompting:** Models receive explicit directives to shape responses, often leveraging templates or structured queries to minimize ambiguity.
- **Example-driven prompting:** This includes few-shot and zero-shot in-context learning, where models are provided with demonstrations or reference structures to dynamically infer task patterns. The study emphasizes that the selection and arrangement of exemplars significantly impact response quality, with retrieval-augmented methods offering dynamic adaptation based on context relevance.
- **Reasoning-based prompting:** Techniques such as Chain-of-Thought (CoT) and its extensions-including Step-Back CoT, Contrastive CoT, and Tree-of-Thought (ToT)-enhance logical structuring by encouraging models to articulate intermediate steps explicitly, improving performance in multi-step problem-solving.

- Self-refinement strategies: These incorporate iterative feedback loops where models critique, verify, and revise their responses. Techniques include:
 - Self-calibration: Estimating confidence in generated outputs.
 - Self-verification: Ensuring logical consistency.
 - Reverse CoT prompting: Reconstructing the original question for validation.
- Critique-based prompting: Models generate counterarguments or assess answer plausibility before refining outputs.
- Editing-based prompting: Structured modifications allow iterative enhancement of generated text.
- Constraint-based prompting: Enforces rules or stylistic boundaries to ensure generated content adheres to predefined linguistic or factual criteria.

The study further evaluates these methods with Prompt sensitivity analysis, that is examining how minor variations in phrasing or token positioning alter model behavior, and robustness testing, assessing output consistency across different prompt formulations.

Lopez-Lira and Tang (2023), “Can ChatGPT Forecast Stock Price Movements? Return Predictability and Large Language Models” investigates the application of large language models in stock prediction, demonstrating that ChatGPT can outperform traditional sentiment analysis methods by interpreting news headlines and financial narratives. The study finds that by leveraging natural language understanding, ChatGPT captures nuanced sentiment signals that are often missed by conventional statistical approaches, leading to improved predictive performance in forecasting stock price movements. A key insight from the research is that ChatGPT’s effectiveness is constrained by the temporal cutoff in its training data, which extends only to September 2021. While this limitation restricts real-time adaptability, it also introduces challenges in back-testing, as any evaluation using data preceding this date risks contamination from information embedded in the model’s pretraining corpus.

A major aspect of the study is the extensive testing of different model types to benchmark ChatGPT’s performance against alternative methodologies. The authors compare the predictive accuracy of ChatGPT-generated sentiment scores with both traditional sentiment analysis models and structured numerical forecasting models. The results indicate that while ChatGPT excels in text-based analysis, particularly in extracting meaningful sentiment from financial news, it struggles significantly with numerical reasoning and temporal understanding. Unlike quantitative models that process structured

financial data, ChatGPT lacks inherent capabilities to perform direct numerical computation, interpret tabular earnings reports effectively, or track time-dependent variables with precision. This limitation leads to inconsistent outputs when handling numerical trends, as the model cannot reliably parse time-series data, compute financial ratios, or forecast earnings growth based on raw numerical inputs.

Despite its weaknesses in numerical and time-based reasoning, the study highlights ChatGPT's exceptional performance in textual sentiment extraction. The model exhibits a strong ability to distinguish between financial language, capturing market sentiment beyond surface-level cues. In sentiment analysis tasks, ChatGPT outperforms traditional lexicon-based and rule-based sentiment models by interpreting contextual relationships within news headlines and press releases, detecting forward-looking statements, and understanding industry-specific terminology. The study reports that ChatGPT's ability to generalize financial sentiment beyond explicit positive or negative terms enables it to better identify market-relevant information than existing lexicon-based models, which struggle with ambiguous or domain-specific phrases.

2.5.3 | Large Language Model (LLM) and Reinforcement Learning (RL)

Pternea et al. (2024), "The RL/LLM Taxonomy Tree: Reviewing Synergies Between Reinforcement Learning and Large Language Models" and Cao et al. (2024), "Survey on Large Language Model-Enhanced Reinforcement Learning: Concept, Taxonomy, and Methods" provide a structured analysis of the integration between Reinforcement Learning (RL) and Large Language Models (LLMs), proposing comprehensive taxonomies that classify their interactions and applications. These studies categorize RL-LLM synergies based on their functional roles and structural relationships, highlighting distinct frameworks where RL and LLMs either fine-tune each other, collaborate in decision-making, or operate in a common environment without direct modification.

Pternea et al. (2024) introduces a three-class taxonomy based on how RL and LLMs interact within a unified system:

- RL4LLM – Reinforcement Learning is applied to fine-tune or enhance LLM performance in natural language processing tasks. This category is further subdivided into:
 - RL4LLM-Fine-Tuning: RL modifies an LLM's parameters to improve task-specific performance. This includes methods using direct human feedback (RLHF) or model-driven tuning without human intervention.

- RL4LLM-Prompt Engineering: RL optimizes LLM prompts iteratively rather than modifying model parameters.
- LLM4RL – LLMs assist RL agents by influencing various aspects of the training framework. This class includes:
 - LLM4RL-Reward Shaping: LLMs design or refine reward functions for RL agents.
 - LLM4RL-Goal Generation: LLMs formulate goals in goal-conditioned RL.
 - LLM4RL-Policy Representation: LLMs directly contribute to policy learning or pretraining.
- RL+LLM – RL agents and LLMs interact in a shared environment without modifying each other’s training processes. This is divided into:
 - RL+LLM-No Language Feedback: The LLM operates independently of RL-driven updates.
 - RL+LLM-With Language Feedback: The LLM dynamically adjusts based on RL agent responses.

Cao et al. (2024) complements this taxonomy by classifying LLM roles in RL-enhanced learning paradigms into four distinct functions as Information Processor – LLMs preprocess and structure input data, reducing complexity for RL agents. This includes state-space encoding and language-to-policy translation. As Reward Designer – LLMs aid in constructing nuanced reward functions, mitigating issues in sparse-reward environments and aligning task objectives. As Decision-Maker – LLMs directly influence action selection by generating policy priors or guiding RL agents through structured task planning. As Generator – LLMs simulate world models, generate synthetic trajectories, and provide interpretable policy rationales.

Both studies emphasize the potential of LLM-augmented RL systems in addressing traditional RL limitations such as sample inefficiency, poor generalization, and the challenge of designing robust reward functions. These taxonomies provide a framework for structuring RL-LLM research, offering a systematic approach to studying their interplay across different learning paradigms.

Yuan et al. (2023), “Skill Reinforcement Learning and Planning for Open-World Long-Horizon Tasks” propose Plan4MC, a hierarchical reinforcement learning architecture designed to tackle the sample inefficiency and task complexity of open-world long-horizon environments, using Minecraft as a testbed. Their approach decomposes

high-level objectives into a sequence of fine-grained, reusable skills-classified as Finding, Manipulation, and Crafting-which are acquired through reinforcement learning with intrinsic rewards.

Central to the system is the novel introduction of a general-purpose Finding-skill, which significantly boosts downstream skill acquisition by pre-conditioning agents with exploratory competence. Rather than relying on LLMs for full trajectory planning, the authors instead use LLMs to construct a static skill graph encoding dependency relations among subtasks. Planning is then executed via a graph-based search algorithm, ensuring stability and correctness in execution. Empirical evaluations over 40 procedurally diverse Minecraft tasks show that Plan4MC achieves strong generalization and substantial performance gains over both imitation- and RL-based baselines, while requiring no demonstrations.

2.5.4 | Investor Risk Profiles

Vakil et al. (2024), "Preferences in Algorithmic Trading: Current Practices and Future Innovations" provide an empirical analysis of how demographic factors influence both the awareness and adoption of algorithmic trading among Indian retail investors. Based on survey data from 124 respondents, the study identifies statistically significant associations between risk-bearing capacity and features such as age, occupation, and annual income. Younger and higher-income investors were more likely to adopt algorithmic strategies, perceiving them as safer and more customizable. These participants also cited reduced human error, enhanced control, and operational convenience as key advantages. The findings point to increased trust in algorithmic trading system, and therefore the need for this system to account for various risk profiles.

Tang et al. (2022), "The Way to Invest: Trading Strategies Based on ARIMA and Investor Personality" introduce investor personality as a structured determinant in optimizing algorithmic trading strategies under varying transaction cost conditions. The authors classify investors into three types-aggressive, intermediate, and conservative. Each profile is subjected to dynamic trading simulations involving gold and Bitcoin, with portfolio allocations informed by ARIMA-based forecasts and optimized using a generalized Sharpe ratio framework.

The study evaluates how transaction fees influence trading behavior across investor types. As transaction costs increase, all investor profiles reduce their trading activity; however, conservative investors exhibit the most pronounced decrease. Notably, variance analysis across fee regimes reveals that conservative investors display the greatest fluctuation in trade counts, reflecting a heightened sensitivity to cost changes. In con-

trast, aggressive investors maintain relatively stable trading frequencies, indicating a lower responsiveness to transactional frictions.

These findings demonstrate that investor personality not only affects baseline risk preferences but also materially shapes how trading behavior adapts under changing cost constraints, supporting the case for algorithmic trading system designed around risk profiles.

2.6 | Conclusion

The primary challenge in algorithmic trading, especially when done with RL algorithms, lies in bridging long-term strategic planning with short-term execution efficiency, grounded in the domain's rational. Achieving superior trading performance requires optimizing order execution while ensuring adherence to overarching investment policies. This research addresses this challenge by integrating LLMs with DRL, using RL+LLM interaction, where the LLM acts as an Information Processor, guiding the RL by synthesizing long term strategies from the environment into the state space of the RL, while the RL does its tactical execution. This setup emulates the top-down hierarchy in an investment bank with the LLM acting as the CIO and the RL as the PM or trading desk.

From the reviewed literature, RL has become a major machine learning method for optimizing executions in algorithmic trading (Dakalbab et al., 2024; Pricope, 2021). Although DDQN may not represent the absolute state-of-the-art in all scenarios, its theoretical clarity, computational efficiency, and suitability as a foundation for advanced extensions, position it as a robust benchmark in RL research (Lazaridis et al., 2020). Consequently, this research utilizes DDQN as the primary RL model for evaluating performance enhancements brought about by integrating LLM-guided decision-making.

For long-term strategic planning, selecting an LLM with robust financial reasoning capabilities, adaptability to dynamic market contexts, and nuanced contextual awareness is critical. Based on the literature reviewed, GPT-4 emerges as the preferred backbone LLM due to its demonstrated excellence in interpreting both structured and unstructured financial data, superior performance in financial decision-making tasks, and proven capacity for generating coherent, contextually-informed strategies (Ding et al., 2024; Lopez-Lira & Tang, 2023; Y. Yang et al., 2020).

The next chapter details the methodology for implementing this LLM-RL framework.

Materials & Methods

This chapter presents the methodology developed to evaluate the integration of LLMs into DRL agents within the domain of quantitative trading. The overarching goal of this integration is to bridge the gap between high-level strategic reasoning and low-level tactical execution in RL, without alterations within the RL's architecture. The proposed architecture reflects top-down workflows found in financial institutions, where investment mandates flow hierarchically from strategic planners to trade desks.

Our approach adopts a hybrid-architecture wherein the LLM agents function as strategic planners, while DRL agents execute the short term trading actions, based on current state, but also in view of the longer term strategic insights provided by the LLMs. To realize this architecture, we developed a pipeline composed of: (i) a structured prompt engineering loops for generating viable strategies, (ii) an RL module conditioned on high-level strategy signals to perform the short-term tactical executions. Collectively, this design enables the system to emulate the reasoning processes of the process of trading ideas and strategies in the industry.

The methodology is structured around two experiments, each aligned with the primary objectives of this research. Together, they evaluate the generative coherence and downstream utility of LLM-guided strategies. Experiment 1 tests whether large language models can produce structured, risk-aware trading policies through prompt engineering alone. Building on those results, Experiment 2 examines whether these strategies can effectively guide reinforcement learning agents across distinct risk profiles. The experimental sequence is thus cumulative: strategy quality is first established in isolation, then assessed within an agentic system tasked with market interaction.

3.1 | Benchmark Research

For our experiments we decided to benchmark our results against the work of Théate and Ernst (2021), previously reviewed in Section 2.5.1, which investigates the application of DRL to algorithmic trading using a DQN algorithm, a method widely adopted in recent literature, as discussed in Section 2.2. Their approach employs a state-of-the-art DDQN architecture specifically designed for single-security trading. The environment is well-defined, with explicit market rules, a curated set of tradable instruments, and a dense reward function that is both simple and effective. Their benchmark spans 30 equities and index-tracking ETFs from U.S., European, and Asian markets between 2012 and 2019, encompassing sectors such as Technology, Finance, Industry, and Agriculture.

Importantly, Théate and Ernst (2021) have made their implementation publicly available on GitHub, facilitating the replication of their results. Their implementation, built in PyTorch, is self-contained, which allows for modular integration of additional components, including the LLM-agents and novel mechanisms introduced in this research.

We successfully replicated their experiments, as shown in Table D.1 in Appendix D.1. Our reproduction yielded results consistent with those reported in the original paper, validating the correctness of our implementation.

This benchmark is particularly well suited to our study for three reasons. First, the authors provide a dense reward function optimized for single-asset trading, enabling transparent comparisons when we evaluate SR improvements. Second, their dataset spans from 2012 to 2019, giving us enough temporal coverage to test our chronological splits. Third, the modular PyTorch code base allows us to graft LLM strategy guidance onto the original DDQN stack without altering network topology, isolating the impact of LLM-derived signals. Where our protocol diverges is in the reduced focus to technological equities.

To ensure fair comparison, we adopt their calculations when reporting the SR and MDD results.

3.2 | Experiments

3.2.1 | Experiments Overview

To evaluate the integration of LLM agents with RL agents for algorithmic trading, this thesis conducts three experiments aligned with the research objectives defined in Section 1.3. Each experiment corresponds to a core component of the proposed hybrid framework.

Experiments were conducted on an NVIDIA GeForce RTX 3050 GPU, with a total training time of approximately 3 hours per security per risk configuration. A consolidated inventory of software versions, seeds, and preprocessing safeguards is provided in Appendix E.

- **Objective 1 - Experiment 1: LLM Trading Strategy Generation**

This experiment evaluates whether an LLM can generate trading strategies that are coherent, economically grounded, and aligned with predefined risk constraints.

- **Objective 2: LLM-Guided RL Agent Performance Across Risk Profiles**

These experiments assess the effect of integrating LLM-generated strategies into RL agents under different trading risk regimes.

- **Experiment 2a: High-Risk Profile** - Tests whether guided agents outperform unguided baselines in terms of cumulative return and SR.
- **Experiment 2b: Low-Risk Profile** - Tests whether guided agents exhibit improved safety, measured by reductions in MDD and Maximum Drawdown Duration (MDDur).

Experiment 1 produces a corpus of validated LLM-generated strategies that are subsequently used to guide agent behavior in Experiments 2a and 2b.

3.2.2 | Experiment 1

This experiment addresses Objective 1 by investigating whether LLM agents, when guided by structured prompt engineering, can generate trading strategies that are logically coherent, economically grounded, and aligned with predefined risk constraints. The central aim is to evaluate the validity and domain relevance of LLM-generated strategies through a process of progressive refinement.

To this end, this experiment incorporates structured prompt tuning, ICL, and iterative feedback loops. These mechanisms are designed to enhance the model's ability to

internalize financial priors, enforce internal logical consistency, and respond adaptively to evaluation signals. Output strategies are subsequently reviewed by domain experts.

This experiment establishes two agents: the strategy agent $\mathcal{A}_{\text{strategist}}$, and the analyst agent $\mathcal{A}_{\text{analyst}}$, for downstream deployment of the strategy π_g , and is foundational for Experiments 2a and 2b.

3.2.2.1 | Data used for LLM Models

For the LLM agents, we constructed a novel multi-modal dataset, aligned with the equities and time frames used in Théate and Ernst (2021), to allow the LLM agents to synthesize trading strategies and evaluate them.

The dataset integrates four canonical categories of financial data, following the taxonomy of López de Prado (2018a): *market data*, *fundamental data*, *analytics*, and *alternative data*. These categories encompass both structured and unstructured modalities, reflecting the heterogeneous information landscape encountered in real-world quantitative trading systems. All features are time-aligned, standardized, and preprocessed prior to integration. Mathematical definitions and transformation formulas are provided in Appendix B.1.

The data pipeline builds upon the benchmark environment of Théate and Ernst (2021). For signal richness and data accessibility, the scope is limited to six large-cap U.S. technology firms: TSLA, AAPL, MSFT, META (formerly FB), GOOGL, and AMZN, selected for their high liquidity, volatility, and prevalence in financial ML literature.

Data were acquired from a combination of public and commercial APIs and are organized as follows:

- **Market data** \mathcal{S}_{mk} : Market data were collected from Interactive Brokers¹. Broader market indicators such as the S&P 500 Index (SPX) and the NASDAQ-100 Index (NDX) were used to contextualize macro-level equity movements. The VIX index, representing expected volatility from S&P 500 options, was also included. Implied Volatility (IV) surfaces were accessed via iVolatility².
- **Fundamental data** \mathcal{S}_{fund} : Firm-level accounting variables were retrieved from SEC EDGAR and SEC-API³, including valuation, solvency, and efficiency metrics. Macroeconomic indicators such as GDP growth, inflation, interest rates, and the

¹<https://www.interactivebrokers.com/api>

²<https://www.ivolatility.com/data-cloud-api/>

³<https://sec-api.io/>

Purchasing Managers’ Index (PMI) were obtained via the FRED API⁴. These data are structured, periodically released, and may be revised retroactively.

- **Analytics** \mathcal{S}_{an} : Technical indicators were computed using TA-Lib⁵ on rolling windows applied to market data. These features encode momentum, volatility, and trend signals.
- **Alternative data** \mathcal{S}_{alt} : Financial news headlines were collected from Alpaca⁶ and transformed into factors using few-shot prompting pipelines based on the LLM-Factor methodology (M. Wang et al., 2024).

Together, these categories of data form the signal universe on which we will run our experiments:

$$\mathcal{SIG}_{uni} = \{\mathcal{S}_{alt}, \mathcal{S}_{fund}, \mathcal{S}_{mk}, \mathcal{S}_{an}\} \quad (3.1)$$

For consistency with the benchmark setup, we adopt the same temporal split as Théate and Ernst (2021): data from 2012–2018 is used for training, and data from 2018–2020 is used for out-of-sample testing. All model artefacts are generated in chronological order with a no embargo between the training and evaluation windows as there is no feature overlap, beyond the intrinsic autoregressive structure of financial time series. Sliding windows advance sequentially without shuffling so that prompts, back-testing runs, and agent training never access observations beyond their effective decision horizon.

Further detail on the data categories mentioned above is provided in the following sub-sections.

Market Data \mathcal{S}_{mk} forms the foundation of short-term financial modeling, providing direct observations of trading behavior, price formation, and liquidity conditions. It encompasses both realized price and volume activity, as well as forward-looking signals inferred from options markets. These features are essential for capturing intraday-to-weekly market dynamics and are widely used in trend detection, volatility modeling, and event-driven trading systems (Dakalbab et al., 2024; Lei & Song, 2024; López de Prado, 2018a; Takara et al., 2023; Théate & Ernst, 2021).

- **OHLCV**: The daily open, high, low, close, and traded volume comprise the fundamental structure of market activity, capturing price dispersion and liquidity across

⁴<https://fred.stlouisfed.org/docs/api/fred/>

⁵<https://ta-lib.org/>

⁶<https://alpaca.markets/docs/api-documentation/>

trading sessions. The Open, High, Low, Close and Volume (OHLCV) signal is represented as a subset $\mathcal{S}_{ohlcv} \subset \mathcal{S}_{mk}$, and is used directly within the RL agent's state space.

- **SPX Index Return:** Represents the daily percentage return of the S&P 500 index, serving as a proxy for overall market direction and beta exposure. This allows the model to contextualize asset-level signals within broader market dynamics.
- **NDX Index Return:** Tracks the daily percentage return of the NASDAQ-100 index, capturing sector-specific movement in technology-oriented equities.
- **Implied Volatility (IV):** Estimated using the Black–Scholes framework (B.27, B.28), implied volatility reflects the market's expectations of future price variance and is derived directly from option prices. From this feature we engineer the Skew Ratio for the perceived asymmetry in risk or directional market sentiment.
- **Historical Volatility (HV):** Calculated from past realized returns over a fixed look-back window (B.29), historical volatility offers a backward-looking measure of actual price variability.
- **VIX Index:** A standardized 30-day forward-looking volatility metric based on S&P 500 option pricing, commonly referred to as the market's "fear gauge".
- **Days to Expiration (DTE):** Represents the remaining time until an option's expiration. DTE plays a significant role in hedging dynamics and can introduce price pressures as expiration approaches.

Fundamental Data \mathcal{S}_{fund} captures the structural financial characteristics of a firm, typically reported in standardized, audited quarterly statements. These indicators reflect long-term economic viability and are generally slow-moving relative to market prices. While fundamental signals are commonly used in portfolio construction and asset screening (Han et al., 2023; H. Yang et al., 2024), their incorporation into tactical strategy generation provides constraints to our agents by serving as low-frequency anchors to the high-frequency Market Data noise.

The following categories of Fundamental indicators have been feature engineered as ratios:

- **Liquidity:** *Quick Ratio, Current Ratio* – Measure short-term solvency by comparing liquid assets to near-term liabilities. Dynamic changes may indicate emerging cash flow constraints or improvements in working capital efficiency.

- **Leverage and Coverage:** *Debt-to-Equity, Debt-to-Assets* – Track shifts in capital structure and leverage exposure. *Interest Coverage* – Reflects the firm’s evolving capacity to service debt obligations.
- **Profitability:** *Gross Margin, Operating Margin, Net Margin* – Capture changes in cost structure and operational efficiency across revenue layers. *Return on Equity (ROE), Return on Assets (ROA)* – Gauge evolving capital productivity and asset utilization.
- **Efficiency:** *Inventory Turnover, Asset Turnover* – Measure changes in the speed at which inventory and assets are converted into revenue.
- **Valuation:** *Price-to-Earnings (P/E), Price-to-Book (P/B), Earnings Yield, EV/EBITDA* – Provide comparative market valuation signals. Temporal trends in these ratios may signal re-rating risk or valuation compression.
- **Growth:** *Revenue Growth, Earnings Growth* – Detects acceleration or deceleration in fundamental performance.

Macroeconomic indicators provide structural context for financial decision-making by describing the aggregate state and directional momentum of the economy. Although typically published at monthly or quarterly frequency, their inclusion in modeling frameworks offers essential information for regime classification (Onozo et al., 2024; Yoon & Fan, 2024; Zhang et al., 2024).

The macroeconomic feature set includes:

- **Gross Domestic Product (GDP):** Captures overall economic output. Changes in GDP are used to detect expansions, contractions, and plateauing regimes.
- **Purchasing Managers’ Index (PMI):** A survey-based leading indicator of industrial activity. QoQ and YoY changes provide early signals of cyclical upturns or slowdowns, particularly relevant for cyclical equities.
- **Producer Price Index (PPI):** Measures upstream inflation. Accelerations in PPI indicate cost pressure and may signal monetary policy responses or margin risk in downstream sectors.
- **Consumer Confidence Index:** Reflects household sentiment regarding income, employment, and general economic conditions. Changes in confidence are useful for anticipating shifts in discretionary spending and retail sector performance.

- **Treasury Yields (10Y):** The 10-year U.S. Treasury yield represents a key long-term interest rate benchmark. Changes in this yield influence asset discounting, portfolio rebalancing, and market-wide risk premia, serving as a foundational reference for evaluating equity valuations and fixed-income instruments.
- **Yield Curve Slope (10Y – 2Y):** The difference between 10-year and 2-year Treasury yields serves as a proxy for expected monetary policy and recession probability. A flattening or inversion is interpreted as a signal of upcoming policy easing or macroeconomic deterioration.

These indicators are commonly used in the financial industry and have been recommended by expert reviewers as core macroeconomic inputs for market analysis and strategy formulation.

Analytics S_{an} and technical indicators provide structured summaries of price action, volatility, and momentum, serving as core features in the LLM-based strategy generation pipeline.

All analytics have been feature-engineered, and are listed below:

- **Simple Moving Averages (SMA):** Computed over 20-, 50-, 100-, and 200-day windows to represent short-, medium-, and long-term trend horizons. These configurations are canonical in both academic literature and trading practice (Chaddha & Yadav, 2022). For each SMA, we compute:
 - *Slope*: Estimated via linear regression across the rolling window, capturing the rate of trend change.
 - *Z-Score*: Measures the standardized deviation from the rolling mean to detect statistically significant price dislocations.
- **Relative Strength Index (RSI):** A bounded momentum oscillator calculated over a 14-day period, commonly used to identify overbought or oversold conditions (Pricope, 2021; Zhang et al., 2024).
- **Average True Range (ATR):** A volatility indicator derived from the 14-day rolling average of the true range. ATR reflects recent price dispersion and is used for volatility-aware position sizing and dynamic stop-loss placement (Dakalbab et al., 2024).
- **MACD and Signal Line:** The Moving Average Convergence Divergence (MACD) is calculated as the difference between the 12-day and 26-day exponential moving

averages, with a 9-day EMA serving as the signal line (Pricope, 2021). From this, we derive:

- *MACD Strength*: The signed deviation of the MACD from its signal line, useful for identifying momentum inflection points.
 - *Slope of MACD Strength*: Captures changes in momentum acceleration, a key input for LLM-based temporal reasoning.
- **Volume-Weighted Average Price (VWAP)**: A cumulative, volume-weighted average that represents the consensus fair value of an asset (López de Prado, 2018a). VWAP is used to as a baseline for intraday and multiday price evaluations.

Alternative Data S_{alt} is the news and the corporate filing data utilized by the analyst agent $\mathcal{A}_{analyst}$.

This study incorporates a structured, prompt-based pipeline to extract latent explanatory factors from financial news headlines, following the *LLMFactor* framework (M. Wang et al., 2024). These features are treated as alternative data, complementing traditional numerical indicators (López de Prado, 2018a; Onozo et al., 2024).

Feature Engineering was performed on the raw data to enable the LLM agents to interpret the complex prompts provided.

In the market data domain, we compute *Put Skew Ratios* (Equation B.26) to quantify directional sentiment as implied by market makers. A put skew greater than one indicates that put options are priced more expensively than their call counterparts, suggesting that market anticipates downside risk. In the case of broad indices, such skew is often interpreted as increased demand for downside protection, signaling systemic uncertainty.

We also provide the LLM with *Weekly Past Returns*, which record the percentage change over the past four weekly intervals. For each trading day t , this includes returns from $t - 5$, $t - 10$, $t - 15$, and $t - 20$ trading days prior. The four-week span was selected empirically to align with the model’s monthly strategy generation frequency.

All fundamental features are expressed as financial ratios, consistent with industry standards for inter-firm comparability. The complete set of formulas is listed in Appendix B.1.

For both fundamental and macroeconomic data, we do not treat raw values as static inputs. Instead, we compute their first-order dynamics using *quarter-over-quarter (QoQ)* and *year-over-year (YoY)* percentage changes. This design enables the model to capture

trends in fundamentals which are likely to be reflected in future price action. Particularly in the case of macroeconomic indicators, we avoid using absolute values, which may implicitly encode temporal anchors (e.g., specific dates or policy regimes) into the prompts, potentially introducing information leakage (see Section 3.2.2.2) since LLMs may memorize specific historical values from their training data if they are back-tested prior to their knowledge cut-off dates (Lopez-Lira & Tang, 2023; Lopez-Lira et al., 2025).

Technical Indicators are computed using TA-Lib⁷, an open-source library for technical analysis. To enable temporal abstraction and reduce high-frequency noise, we apply a 20-day rolling window, selected empirically to match the cadence of monthly strategy generation. These indicators are further augmented with slope and z-score derivatives, allowing the LLM to reason about directional trends and magnitude shifts. While some recent methods opt for natural language representations of these signals (M. Wang et al., 2024; Zhang et al., 2024), we found that purely statistical feature engineering yields superior empirical performance in the context of automated strategy generation (Lopez-Lira & Tang, 2023). See Appendix B.1 for the formulation of all technical indicators.

For news data, we process it into a *News Factor* representation, wherein an LLM generates 2–5 interpretable latent factors per summarized news item. Each factor captures a market narrative, such as regulatory pressure or supply chain disruption, serving as a proxy for the underlying narrative context. To prevent forward-looking bias in news-derived features (see Section 3.2.2.2), we anonymize all named entities and temporal references using in-context substitutions (e.g., “Tesla” becomes “the COMPANY”) to prevent recall bias before the knowledge cutoff date (Lopez-Lira et al., 2025).

3.2.2.2 | LLM Selection and Back-Testing

The selected LLM for this experiment is GPT-4o Mini, chosen based on performance in financial reasoning or general tasks, for reference see Section 2.6. Prior research consistently identifies GPT-4 class models as top-performing in financial contexts, demonstrating superior risk-adjusted returns across tasks involving market understanding and structured inference (J. Huang & Chang, 2023; Lopez-Lira & Tang, 2023; Yu et al., 2024).

The GPT-4o Mini variant was selected to balance computational efficiency and model capacity. It offers a 128,000-token context window, a maximum prompt size of 16,384 tokens, and a competitive cost structure of \$0.15 per million input tokens, enabling scalable deployment of large prompt templates without incurring prohibitive costs.

Despite their advantages, a key limitation of LLMs in financial domains is their static knowledge base. As noted by Yu et al. (2024), LLMs are unable to dynamically update

⁷<https://ta-lib.org/>

their internal representations and must instead rely on iterative question-answering to simulate adaptation. This limitation also introduces risks such as data leakage and hindsight bias, particularly when real-world entities or timestamps can be recognized from the model’s training corpus.

To mitigate this, all prompts were anonymized to remove identifiable references. Security names, employee identities (e.g., “Elon Musk”), and calendar years (e.g., “2020”) were abstracted to neutral placeholders (e.g., “The CEO”, “The Company”, “Next Year”). This anonymization strategy allowed the model to exploit structural signals such as news factors or price patterns, without using memorized knowledge of specific world events. Table 3.1 presents the knowledge cutoff dates for major LLMs, confirming the alignment between the GPT-4o Mini model and the benchmark dataset used in this experiment.

Given that an LLM’s knowledge cutoff determines the temporal boundary of its internal representation of the world, which is directly relevant to the experiment. In our case, the model’s historical training window substantially influences its reasoning context. For instance, an LLM such as GPT-3.5, with a cutoff in January 2022, may be unable to reason effectively or draw analogies within market regimes that emerged afterward, such as the post-COVID industry landscape or the AI-driven market behaviors observed in 2025 and beyond. This limitation can affect the validity of trading experiments conducted far beyond the model’s training horizon, as the model’s priors may no longer align with the prevailing market conditions.

Table 3.1: Knowledge Cutoff Dates for GPT (OpenAI, 2024)

LLMs	Cutoff Date	Provider
GPT-4o	October 2023	OpenAI
GPT-4	April 2023	OpenAI
GPT-3.5	January 2022	OpenAI

The cutoff dates in Table 3.1 align the GPT-4o Mini deployment with the 2012–2020 market history but introduce potential recall bias. Strategies inferred from pre-2023 knowledge may overweight growth stocks given the historic bull-run post the Great Financial Crisis in 2008. To mitigate this, we employed anonymization techniques, ensuring that the model could not leverage memorized events or data points. This approach allowed the LLM to focus on structural patterns and relationships within the provided data, rather than relying on specific historical knowledge.

3.2.2.3 | Baseline Prompt

We began by constructing a baseline prompt, referred to as Prompt 0, to serve as the control condition for evaluating the impact of prompt engineering techniques.

To isolate the effect of prompt structure on the quality of LLM-generated trading strategies, we compare a structured prompts against this minimal baseline. The baseline prompt consists of a simple instruction to generate a trading strategy based solely on the provided price data and technical indicators, without any additional context or guidance.

The baseline prompt, detailed in Appendix C.1, includes only raw price inputs and standard technical indicators: Simple Moving Average (SMA)-20/50/100/200, Relative Strength Index (RSI), and Moving Average Convergence Divergence (MACD). It omits all structured reasoning and in-context learning. This light-weight configuration mirrors practices found in the algorithmic trading literature (Chaddha & Yadav, 2022; Dakalbab et al., 2024; Liu, Yang, et al., 2022; Takara et al., 2023) and reflects heuristics commonly employed by amateur and intermediate traders (Alsini et al., 2024). Its validity as a baseline was corroborated through expert consultation, and it served to set the candidate prompt (Prompt 1) and V^* target for the regret–minimization loop.

3.2.2.4 | Prompt Refinement Process

Prompt refinement builds on recent works in financial prompt engineering (Dai et al., 2023a; Schulhoff et al., 2024; Yu et al., 2024). All prompt variants are evaluated using both intrinsic model metrics: Perplexity (PPL), H_{LLM} and extrinsic trading metrics: the backtested SR. The iterative tuning process is illustrated in Figure 3.1.

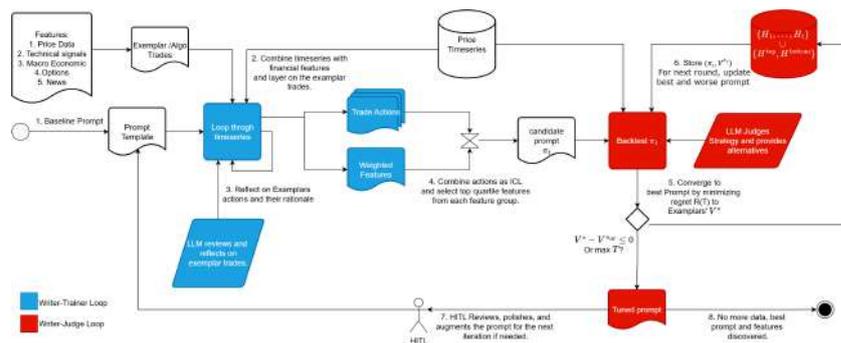


Figure 3.1: Prompt Tuning Workflow

The first stage of refinement involved constructing an overcomplete prototype referred to as the *writer-trainer prompt*, inspired by S. Wang et al. (2024) and detailed in

Appendix C.2. This expanded version incorporated a broad array of structured financial indicators, deliberately excluding unstructured textual data for speed, while retaining complete coverage of the universe of quantitative signals. It functioned as a high-coverage scaffold, from which more targeted prompt variants could be derived through ablation and simplification.

The design of this pipeline followed the guidance of data-mined trading rule discovery outlined in Vanstone and Hahn (2015), exhausting most of the inputs prior to tuning or specializing the model. Ding et al. (2024) also recognizes that LLMs can act as trading pattern miners (referred to as Alpha). To further condition the model on domain behavior, exemplar trades were appended to the historical input series (Step 1, Figure 3.1). These exemplars served to seed priors in the LLM and were drawn from two sources: a heuristic expert trading algorithm (Appendix 3) and suggestions from Human-in-the-Loop (HITL).

This exercise was performed on two instruments TSLA and AAPL, chosen for their characteristics and volatility profiles, exposing the LLM to a heterogeneous set of decision contexts. Trade examples were sampled from the 2012–2020 period, ensuring the LLM sees a wide spectrum of market conditions and trades.

The LLM was then instructed, using this prompt, to engage in structured reflection over each exemplar (Step 2, Figure 3.1). Specifically, it was asked to:

- Justify each trade in terms of underlying causal reasoning aligned with the stated portfolio objective and risk tolerance. For example, in risk-seeking scenarios, the model was expected to highlight overbought signals such as `Stock_Data.RSI` exceeding 70.
- Rank features within each feature group according to their inferred contribution to the strategy. These rankings employed a Likert-style schema (1 = minimal, 2 = neutral, 3 = critical) to formalize feature prioritization (Gonen et al., 2023; Schulhoff et al., 2024).

This methodology follows reflective prompting techniques (Schulhoff et al., 2024; Yu et al., 2024), where the LLM is guided to construct modular reasoning chains that connect external heterogeneous signals to actionable outputs. For instance, a long position might be justified by the combination of upward price momentum, flattening yield curves, and reducing option put-skew.

The universe of features in the *writer-trainer prompt* were reduced to those that consistently ranked as critical, while less impactful indicators were removed.

The top quartile of features from each domain: technical, macroeconomic, fundamentals, and options, was then aggregated to compose a refined prompt template, also detailed in Appendix C.3. This template served as the basis for the next stage of generation, illustrated in Step 3 of Figure 3.1.

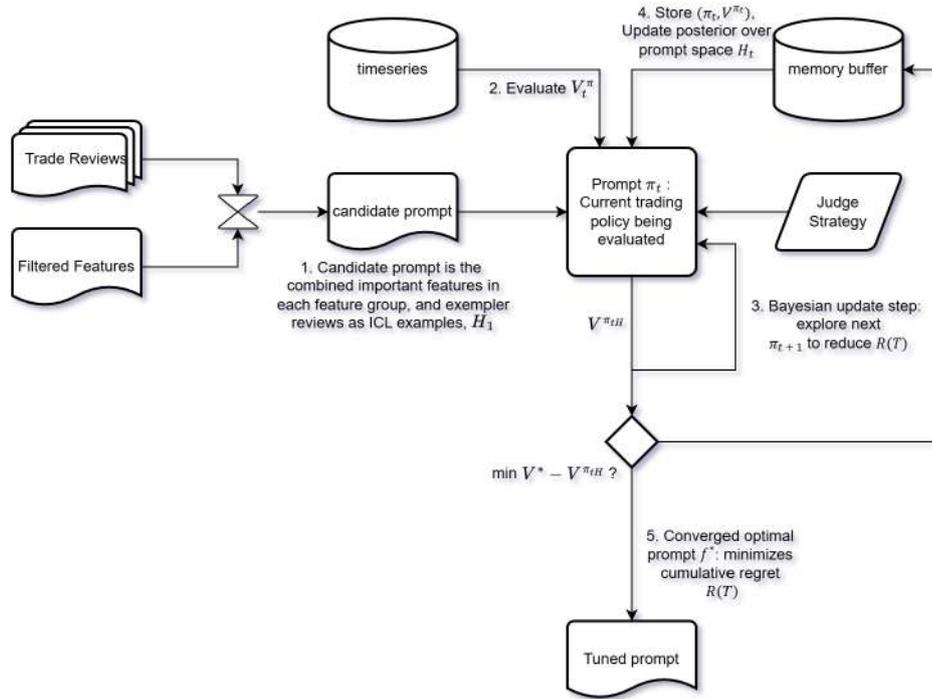


Figure 3.2: Prompt Refinement Loop with Writer-Judge (S. Wang et al., 2024)

Following prompt assembly, a structured refinement cycle was implemented using the writer-judge framework adapted from S. Wang et al. (2024). As shown in Figure 3.2, this framework orchestrates a loop between candidate prompt generation (Writer), backtest-based evaluation, and structured assessment (Judge), with each iteration informing the posterior over prompt space.

It proceeds in five stages:

- 1. Prompt Assembly:** At each round t , a candidate prompt π_t is constructed by combining top-quartile features from prior high-performing strategies. These features span categories such as technical indicators, macroeconomic variables, fundamentals, and options. Prompts also include in-context examples H_1 drawn from a memory buffer of past completions.

2. **Strategy Evaluation:** The prompt π_t is applied to a sampled historical time window to generate a trading strategy via the LLM. The resulting strategy is backtested, and its performance is recorded as V^{π_t} , measured using the SR in alignment with our evaluation criteria.
3. **Posterior Update:** The prompt-performance pair (π_t, V^{π_t}) is added to memory. The next prompt π_{t+1} is formed by mutating a top-performing prompt and applying edits suggested by the Judge component.
4. **Regret Minimization:** The goal is to reduce cumulative Bayesian regret:

$$\mathcal{R}(T) = \mathbb{E} \left[\sum_{t=1}^T (V^* - V^{\pi_t}) \mid \mathcal{H}_t \right], \quad (3.2)$$

where V^* is the return of the (unknown) optimal strategy, V^{π_t} is the current policy, and \mathcal{H}_t is the memory buffer containing the previous policies, Judge feedback, and their results. As the refinement proceeds, expected regret should decrease as prompts converge toward higher-value regions of the search space.

5. **Convergence:** If regret falls below a predefined threshold, the corresponding prompt π^* is selected as the final strategy-generation template. If no such prompt is found within T rounds, the best-performing prompt so far is returned.

The *Judge* (S. Wang et al., 2024) detailed in Appendix C.4, is a secondary LLM prompted to evaluate the logical structure, feature usage, and alignment of each strategy with financial objectives. When flaws are detected, the Judge provides corrections and ICL exemplars, which are incorporated into the next prompt iteration.

This iterative cycle enabled refinement of both the structure and content of the prompt, resulting in a final candidate f^* with empirically minimized regret $\mathcal{R}(T)$. Adaptive modifications made to the prompt during this process included:

- Integration of the Put Skew, when high-momentum signals co-occurred with rising a rising VIX, enhancing downside risk detection.
- Down-weighting of the RSI technical signal during volatile market periods, due to repeated misalignment with actual reversals in the backtested sample.

This process produced a tuned baseline prompts (see Appendix C.5), generalized across risk profiles. Each was validated on the benchmark’s out-of-sample test set using both SR and PPL as evaluation metrics.

3.2.2.5 | Prompt Improvement 1 - In-Context Memory

The first discretionary enhancement to the prompt design introduced an In-Context Memory (ICM), resulting in a revised prompt format documented in Appendix C.6. Inspired by the FinMEM framework (Yu et al., 2024), this enhancement embeds prior strategy outputs and performance diagnostics directly into the prompt, enabling experience-based adaptation over time. The approach parallels the Writer–Judge memory loop introduced in Section 3.2.2.3, but extends it by explicitly integrating model-generated reflections.

The ICM mechanism allows the LLM to condition its current decision on past strategy rationale and observed performance. Specifically, when available, the prompt includes a variable ‘Last_Strategy_Used_Data’, allowing the model to evaluate mismatches between prior expectations and actual outcomes. This design encourages the LLM to identify underperforming components—such as overreliance on volatility or misaligned macro signals—and update its reasoning accordingly.

Feature attribution is made explicit through a structured Likert-style weighting system. Each input signal used in the prior strategy is assigned an importance score from 1 to 3:

- **1** – Minimal impact; may be ignored in future iterations.
- **2** – Moderate contribution; relevant but non-critical.
- **3** – High relevance; central to the strategy’s performance.

These weights are reused in subsequent prompt generations to guide the model’s feature prioritization. This enables the LLM to shift its attention toward or away from signals based on historical efficacy—without requiring retraining or parameter updates (Schulhoff et al., 2024).

The prompt also distinguishes between high-risk and low-risk configurations, adjusting its interpretation of indicators accordingly. For example:

- **High-Risk:** Emphasizes momentum (e.g., RSI, MACD) and volatility-based signals (e.g., ATR, ATM IV).
- **Low-Risk:** Prioritizes stability metrics (e.g., Debt-to-Equity, Current Ratio) and downside risk mitigation (e.g., ITM IV skew, Treasury yields).

A final reflection step encourages the model to re-evaluate its own logic using Chain of Thought (CoT) style reasoning. If ‘Last_Strategy_Used_Data’ is not ‘null’, the model

is instructed to compare current assumptions with prior outcomes and adjust signal weightings accordingly.

Together, these components transform the prompt from a static instruction template into a dynamic, memory-informed reasoning system. This iterative structure supports cumulative learning and self-correction across multiple rounds of strategy generation, enhancing both robustness and adaptability of outputs over time.

3.2.2.6 | Prompt Improvement 2 - Instruction Decomposition

The second refinement phase focused on explicitly structuring the strategy-generation prompt using modular decomposition and CoT reasoning (M. Wang et al., 2024; Zhang et al., 2024). This approach, implemented in the revised prompt template (Appendix C.7), aimed to improve modularity, and reasoning traceability in the LLM's outputs.

The system context was decomposed into six discrete modules: stock analysis, technical analysis, fundamental analysis, macroeconomic analysis, options analysis, and a final reflection step, similar to the setup in Zhang et al. (2024). Each module contains a structured breakdown of relevant features, decision heuristics, and example completions derived from prior high-performing strategies. This decomposition guides the LLM to reason over each domain independently before synthesizing an integrated directional action (LONG or SHORT).

Within each module, domain-specific prompts incorporate few-shot ICL exemplars to anchor the model's interpretation of scenario-relevant patterns. For example:

"Close price above 20MA and 50MA with steep 20MA slope signals bullish momentum. Weekly returns confirm a sustainable uptrend."

These analogical examples act as inductive biases, helping the model match current market conditions with historically successful signal configurations.

Decision logic is structured around known trading heuristics, or those discovered in the *writer-judge* loop. For example:

- **Macro analysis** uses slope conditions on SPX and VIX to identify risk-on/off regimes, complemented by PMI and GDP trends to infer economic cycles.
- **Fundamental analysis** balances growth and stability via profitability ratios, earnings growth, and leverage metrics.
- **Options analysis** captures sentiment and skew dynamics by comparing ATM/OTM/ITM implied volatilities.

The final reflection module promotes dynamic adjustment by encouraging the model to review the last strategy’s performance (if available) and weigh feature importance using a Likert-style scale (1–3) (Schulhoff et al., 2024). This supports backward reasoning and facilitates more nuanced CoT adaptation in the presence of shifting market behavior.

By enforcing instruction decomposition with ICL and CoT logic within each component, the prompt enhances model outputs and also simplifies human review and domain expert validation.

At this point, the strategy agent \mathcal{A}_s , has been trained prompted to create the best possible strategy π_g that can be synthesized to down-stream agents.

3.2.2.7 | Prompt Improvement 3 - News Factors

The third and final refinement introduced unstructured textual data into the strategy generation process. This extension, presented in Appendix C.8, augmented the LLM’s input space with firm-specific news summaries and corporate filings, leveraging its native language understanding capabilities through the use of an additional agent: the analyst agent \mathcal{A}_a

The refinement builds on the factor induction framework proposed by M. Wang et al. (2024), wherein latent factors are extracted from financial texts using prompt-driven reasoning. In this work, the same principle was applied: headlines and short passages were incorporated into the prompt, and the LLM was tasked with mapping these to relevant trading implications via CoT prompting. To ensure consistent grounding, all extracted insights were aligned with predefined signal groups (technical, macro, fundamentals, options) and contextualized accordingly.

To prevent leakage from memorized world events, an anonymization protocol was adopted in the prompt. All entity names, company identifiers, and absolute calendar dates were abstracted into neutral placeholders. For example, the sentence:

“Tesla CEO Elon Musk sets goal of making 500,000 cars in 2019”

was rewritten as:

“The COMPANY’s CEO sets goal of making 500,000 cars next YEAR.”

This transformation preserved the causal and structural implications of the statement (e.g., bullish forward guidance, high-capex growth signal) while preventing the model from anchoring on real-world historical knowledge.

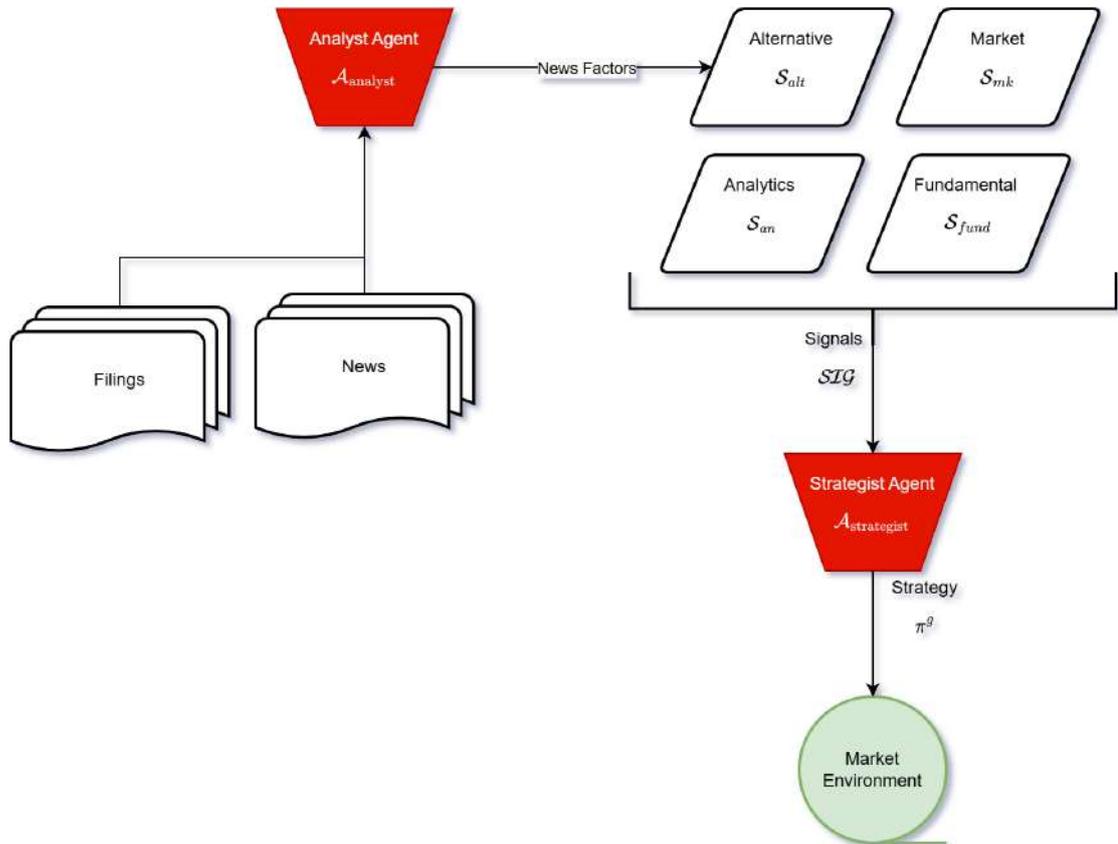


Figure 3.3: LLM Agents Architecture.

News-based features were treated as an additional feature group in the final prompt architecture. They were weighted, like structured indicators, via Likert-style scoring based on inferred importance during prior prompt iterations. Their inclusion allowed the model to reason over forward-looking qualitative information, expanding beyond the confines of purely numerical signal interpretation.

Overall, this refinement enabled the model to integrate heterogeneous modalities, numerical indicators and unstructured text, into a unified strategic rationale. The resulting prompt architecture supported a multi-layered decision framework that reflected realistic investor reasoning and offered a more robust testing ground for evaluating the LLM's ability to align diverse data sources in pursuit of profitable trading strategies.

Figure 3.3 details the initial architecture and data pipelines that go into the LLM agents in synthesizing the strategy, or global policy π_g .

3.2.2.8 | Parameters

To promote reasoning diversity and discourage repetitive outputs from prior completions in the memory buffer, prompt generation was conducted under the following hyperparameter settings:

- **Temperature:** Set to 0.7 to encourage diversity in generated reasoning paths and reduce determinism (Lopez-Lira & Tang, 2023; Yu et al., 2024).
- **Frequency Penalty:** Set to 1.0 to penalize repetition in token generation, promoting lexical variability and reducing overused phrases.
- **Presence Penalty:** Set to 0.25 to incentivize the inclusion of novel reasoning patterns and feature combinations across prompt generations.
- **Regret Threshold:** *Writer-Judge* convergence is defined when the expected regret $\mathcal{R}(T) < 0.15$, where $T = 3$ is the maximum number of refinement iterations.
- **Initial Sharpe Ratio V^* :** Defined as 0.10, and updated which each policy that yields a SR greater than this value.
- **Rolling Window Size:** Set to 20 trading days for all statistical transformations (e.g., slope, z-score).
- **Technical Indicator Parameters:** Default configuration is used for all indicators (e.g., 14-day RSI, 12/26/9 MACD).

All hyperparameters except *Temperature* were selected through empirical testing.

For test runs, a new strategy was generated on a monthly basis. This cadence was chosen for two reasons: First, as described in Section 1.4, the architecture emulates organizational hierarchies common in financial institutions, where portfolio managers typically receive updated guidance on a monthly or quarterly basis unless significant market events occur. Second, generating high-frequency strategies is cost-prohibitive at scale. Each strategy costs approximately \$0.002 to \$0.014 to generate, making sub-monthly inference over 8 years of back-testing records and multiple tickers economically inefficient.

3.2.2.9 | Quantitative Evaluation Criteria

The evaluation of LLM-generated trading strategies is based on three primary quantitative metrics: the Sharpe Ratio (SR), Perplexity (PPL), and token-level entropy (H_{LLM}).

These metrics jointly capture financial performance, model confidence, and generation-time uncertainty.

- **Sharpe Ratio (SR)** - Quantifies the risk-adjusted return of a trading strategy. It is used here as a proxy for domain alignment, with higher values suggesting that the generated strategies successfully optimize returns relative to volatility. Following prior studies such as Lopez-Lira and Tang (2023) and Yu et al. (2024), we interpret elevated SR as indicative of improved reasoning in financial decision contexts.
- **Perplexity (PPL)** - Measures the model’s confidence in its generated output. It is defined as the exponentiated negative average log-likelihood over the generated sequence:

$$\text{PPL} = \exp \left(-\frac{1}{N} \sum_{t=1}^N \log p(w_t | w_{<t}) \right), \quad (3.3)$$

where N is the number of generated tokens and $p(w_t | w_{<t})$ is the conditional probability of token w_t . We report PPL over the entire generated output, which includes both the predicted trading action (e.g., LONG or SHORT) and the accompanying rationale. Lower PPL indicates that the model assigns higher likelihood to its own generation, making it a useful diagnostic metric for prompt effectiveness. Our methodology is inspired by Gonen et al. (2023), who show that lower perplexity is correlated with better performance on the downstream task. While their study measures PPL on the prompt tokens, this is no longer feasible with the GPT4 model used. Consequently, we compute PPL solely on the completion.

- **Entropy H_{LLM}** - Reflects the model’s uncertainty at each generation step, calculated using a truncated token-level entropy over the top- k probabilities exposed by the API:

$$H_{\text{LLM}} = \frac{1}{N} \sum_{t=1}^N \left(\sum_{v \in V_k} -p_t(v) \log p_t(v) - p_{\text{tail},t} \log p_{\text{tail},t} \right), \quad (3.4)$$

where V_k is the top- k token set at step t , $p_t(v)$ is the model-assigned probability for token v , and $p_{\text{tail},t} = 1 - \sum_{v \in V_k} p_t(v)$ is the residual mass of unobserved tokens. This yields a conservative lower bound on entropy, following the approximation proposed in Kaltchenko (2025). Unlike PPL, which measures confidence in the chosen output sequence, H_{LLM} captures the breadth of alternatives the LLM considered, and is thus a useful diagnostic for the model’s decisiveness.

Taken together, PPL and H_{LLM} provide a two perspectives on generation behavior, distinguishing between fluency and confidence in strategy output:

Profile	Interpretation
Low PPL, Low H_{LLM}	Fluent and Confident; indicative of a well-aligned generation.
Low PPL, High H_{LLM}	Coherent but hesitant; the model deliberated among multiple completions.
High PPL, Low H_{LLM}	Confident but potentially misaligned to the prompt; generated low-quality or incorrect output.
High PPL, High H_{LLM}	Incoherent and Indecisive; suggests full misalignment or generation failure.

This classification supports a nuanced evaluation of model behavior across prompt variants and risk profiles, aiding both interpretability and debugging of strategy generation quality.

3.2.2.10 | Qualitative Evaluation Criteria

- **Expert Review Score (ERS)** - a human-grounded evaluation score derived from expert feedback. Following a rubric, experts rate the rationale quality of LLM-generated strategies across three dimensions: logical soundness, domain fidelity, and risk handling. Each dimension is scored on a 3-point ordinal scale {1, 2, 3} for poor, average and good. The rubric is detailed in Table 3.2.

For qualitative review we adopted the approach of Demajo et al. (2020), which employed seven domain experts and a group of laypeople to examine contrastive insights. We selected ten participants for the expert review: five finance professionals in senior positions, and five finance amateurs (or that do not execute trades if they work in the domain). The rationale behind having retail traders is that the LLM training data is what they would be reading to prepare for their trades, and would produce contrastive insights across experience levels (Demajo et al., 2020). Each expert reviewer received anonymized data for three financial instrument over a one-year horizon, stripped of identifiers such as asset names and timestamps to prevent lookahead bias.

Each dataset included:

1. Historical technical indicators.
2. Fundamental valuation metrics.

3. Macroeconomic indicators.
4. Condensed and anonymized firm-specific and macroeconomic news headlines.

Review survey was benchmarked at around 15 to complete, followed by a 60 minute discussion (Demajo et al., 2020). The survey included an initial directional prediction for the reviewer, followed by the LLM rationale review, and rubric scoring (Hoffman et al., 2018). Prior to viewing the LLM rationale, reviewers submitted their own directional trade (LONG or SHORT) and justification, this was done to activate their domain-relevant mental model (Hoffman et al., 2018). Afterwards, participants reviewed the LLM-generated rationale and evaluated it using the rubric (See Table 3.2).

The following discussion of up to 60 minutes, allowed us to understand the rationale and criticism of the reviewer, acquire their domain knowledge, curate the data or add new sources, edit the prompts and gather exemplars for the final prompt.

Criterion	Score -1 (Poor)	Score 0 (Average)	Score 1 (Good)
Rationale	Contradicting and/or with logical errors.	Mostly coherent, with some gaps.	Logical and valid reasoning.
Fidelity	Not applicable to real-world markets.	Somewhat applicable, but unrealistic.	Performs like a trained professional.
Safety	Ignored the risks, or confabulated.	Recognized the risks but did not act.	Recognized and mitigated the risks.

Table 3.2: Rubric for evaluating LLM-generated trading rationales.

For the published survey, see Appendix F. Note that while some surveys originally used a scale from -1 to 1, the scores were all normalized to a 1 to 3 range for consistency.

3.2.3 | Experiment 2

This experiment tests whether a fixed RL architecture, when guided by strategies generated from a risk-aware LLM, can align with both high-risk and low-risk investor profiles. Crucially, the underlying RL agent’s architecture remains unchanged across conditions; all behavioral modulation arises from the LLM’s adaptation to risk-specific prompts.

The core hypothesis is that access to structured, economically grounded strategies enables the RL agent to make more informed decisions, leading to improved outcomes in both high-risk and low-risk regimes. To test this, Experiment 2 is divided into two sub-experiments: Experiment 2a and Experiment 2b.

Together, these experiments evaluate the functional utility of LLM-based strategy guidance in dynamic trading environments.

3.2.3.1 | Data used for RL Model

For LLM data, we apply the same data used in Experiment 1.

For the RL data, we replicated the same feature set used Théate and Ernst (2021) for the RL agent’s state space, given that we will use their DDQN architecture.

The RL agent receives a state representation comprising a rolling 30-day window (Théate & Ernst, 2021) of the raw market data \mathcal{S}_{ohlcv} .

Feature Engineering Building upon this foundation, we extend the state space with additional features derived from the LLM agent’s proposed strategy, conceptualized as the global guidance policy π_g provided to the RL agent:

- **Signal Direction** ($\text{dir}(\pi_g)$): A binary variable indicating the directional stance of the strategy, where 0 corresponds to a *Short* recommendation and 1 to a *Long* recommendation.
- **Signal Strength** ($\text{str}(\pi_g)$): A continuous variable capturing the product of strategy’s confidence and the LLM’s implicit model certainty derived from its entropy.

Along with the trade action, each strategy π_g returns a discrete confidence score based on a novel 3-point Likert-style rubric. Higher scores indicate stronger semantic alignment with domain-relevant features and greater evaluator confidence:

- **3 (High Confidence)**: Strong convergence across key features; the recommendation is well-supported by coherent, corroborative signals. This score is appropriate when the strategy exhibits high-risk/high-reward characteristics under a *High-Risk* profile, or demonstrates robust low-risk behavior under a *Low-Risk* profile.
- **2 (Moderate Confidence)**: Partial alignment with moderate evidential support; some ambiguity or uncertainty remains. Assigned when the strategy partially matches the intended risk configuration but may lack supporting confirmations or show marginal inconsistencies.
- **1 (Low Confidence)**: Weak alignment or internal contradictions; justification is insufficient or signals are inconsistent. Typically applies when the strategy contradicts the assigned risk profile-e.g., when a *Low-Risk* profile is paired with high

downside volatility, or a *High-Risk* profile yields low-reward conservative behavior.

The discrete confidence score is normalized to a continuous value $\mu_{\text{conf}} \in \{0.33, 0.67, 1.0\}$ via:

$$\mu_{\text{conf}} = \frac{\text{Likert}}{3} \quad (3.5)$$

To account for the uncertainty inherent in the LLM’s output, we define an entropy-adjusted certainty score:

$$C = \varepsilon + (1 - \varepsilon)(1 - H) \quad (3.6)$$

where $H \in [0, 1]$ denotes the normalized Shannon entropy of the LLM’s output distribution (see Section 3.2.2.9), and $\varepsilon > 0$ enforces a non-zero lower bound on certainty. This formulation stabilizes downstream confidence propagation and ensures that even high-entropy completions retain minimal influence.

The final signal strength associated with the LLM-guided strategy π_g is computed as:

$$\text{str}(\pi_g) = \mu_{\text{conf}} \cdot C \quad (3.7)$$

where $\mu_{\text{conf}} \in [0, 1]$ denotes the normalized evaluator confidence score derived from expert Likert ratings. The signal strength $\text{str}(\pi_g)$ integrates semantic alignment with statistical certainty, functioning as a principled filter to modulate RL agent behavior.

This scoring method is consistent with the recommendations of Yona et al. (2024), who emphasize the need for evaluative functions that jointly capture quality of reasoning and internal model confidence. High values of $\text{str}(\pi_g)$ indicate coherent and reliable strategies, while lower values highlight cases of ambiguity and the model’s uncertainty.

3.2.3.2 | Overview of LLM/RL Hybrid Architecture

The foundation of our system builds upon the DDQN architecture, environment, and training procedures described in Théate and Ernst (2021), their architecture will be our RL agent \mathcal{A}_{rl} . On top of this baseline, we integrate our LLM-based agents $\mathcal{A}_{\text{strategist}}$ and $\mathcal{A}_{\text{analyst}}$, developed in Experiment 1, as illustrated in Figure 3.4.

All numerical features are normalized using dynamic min–max scaling over a rolling window. The resulting input is flattened into a fixed-length vector and passed to the DDQN for action evaluation.

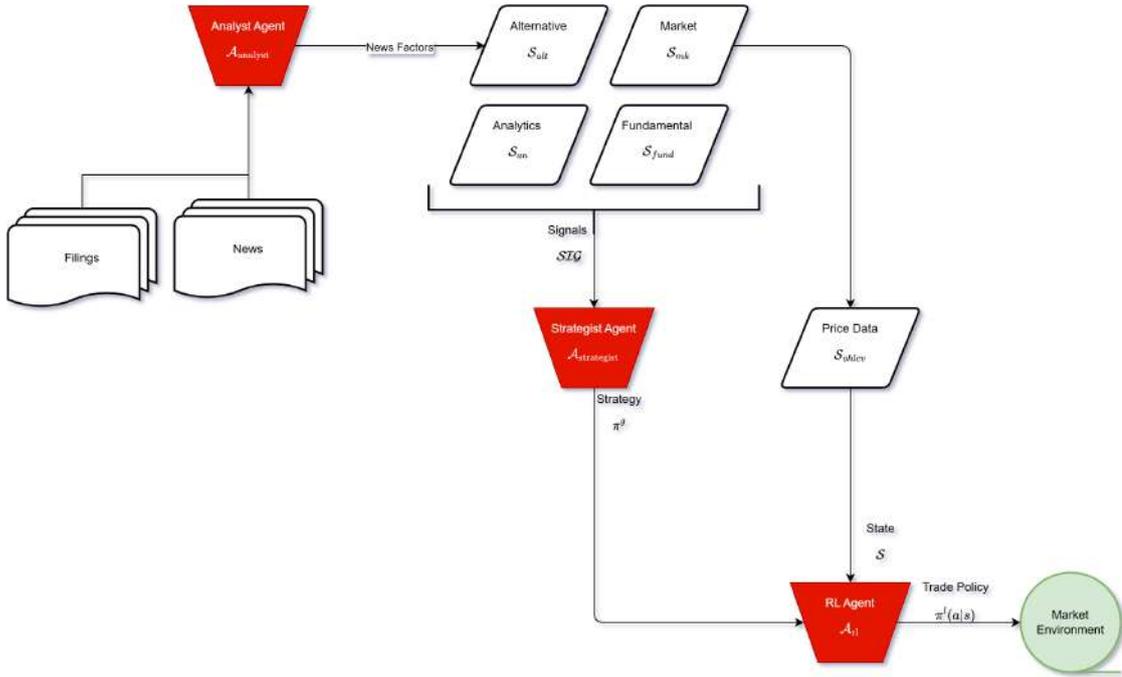


Figure 3.4: LLM/RL Hybrid Architecture.

While we retain the training loop outlined in the original paper, we aimed to reduce our inference overhead by precomputing the LLMs-outputs once per instrument and risk profile, these outputs remain fixed throughout the training cycle. The cached policy guidance π_g is concatenated with the agent’s state space and reused across all episodes, significantly reducing computational costs while preserving global policy alignment.

3.2.3.3 | Parameters

To ensure comparability with prior work, we adopt the hyperparameter configuration established by Théate and Ernst (2021).

For the LLM agents $\mathcal{A}_{\text{strategist}}$ and $\mathcal{A}_{\text{analyst}}$, we retain the same settings described in Experiment 1 (Section 3.2.2.8).

3.2.3.4 | Evaluation Criteria for Experiment 2

To assess the performance of LLM-guided RL agents under distinct risk profiles, we define tailored evaluation criteria corresponding to high-risk (Objective 2a) and low-risk (Objective 2b) configurations. All results are reported on held-out data from the benchmark dataset (2018–2019), processed using the same pipeline described in Théate and Ernst (2021). Each agent is trained and evaluated across 25 independent replications,

with 50 episodes per financial instrument, to ensure statistical robustness. Experiments were conducted on an NVIDIA GeForce RTX 3050 GPU, with a total training time of approximately 3 hours per security per risk configuration.

High-Risk Profile (Experiment 2a) In the high-risk configuration, the objective is to evaluate whether LLM-guided agents achieve superior performance relative to the benchmark (Théate & Ernst, 2021). The primary focus is on risk-adjusted return generation:

- **Sharpe Ratio (SR)** - Measures the ratio of excess return to volatility. This ensures that elevated returns are not merely the result of increased risk exposure but reflect efficient risk-adjusted performance.

Low-Risk Profile (Experiment 2b) In the low-risk configuration, the emphasis shifts to preserving capital and ensuring longevity. The aim is to assess whether LLM-guided agents effectively reduce exposure to losses relative to the benchmark (Théate & Ernst, 2021):

- **Maximum Drawdown (MDD)** - Captures the largest peak-to-trough loss over the evaluation period. A lower MDD indicates stronger defensive behavior and improved resilience under adverse market conditions.

The benchmark thresholds used for evaluation are reported in Table 3.3.

Table 3.3: Benchmark thresholds.

Instrument	SR	MDD
TSLA	0.62	0.65
META	0.15	0.45
AMZN	0.42	0.19
GOOGL	0.23	0.25
MSFT	0.99	0.17
AAPL	1.42	0.45
Mean	0.64	0.36

3.3 | Conclusion

This chapter presented the methodological framework developed to evaluate the integration of LLM agents with RL agents in the context of algorithmic trading. The core objective was to assess whether strategic reasoning, as modeled by LLMs, could enhance the tactical behavior of RL agents across varying risk profiles-emulating the layered decision-making structures found in institutional finance.

To this end, a multimodal dataset was constructed, combining structured financial indicators with unstructured news-derived signals. These were normalized, feature-engineered, and embedded into prompt templates to support risk-sensitive strategy generation via LLM agents.

The methodology comprised three interlinked experiments. Experiment 1 evaluated whether LLMs could generate economically coherent and risk-aligned strategies through iterative prompt engineering. Experiment 2 was divided into two subparts: Experiment 2a assessed the performance of LLM-guided RL agents under a high-risk mandate, focusing on SR; Experiment 2b examined a low-risk configuration, evaluating improvements in MDD. Both subparts were tested on the same RL architecture without alterations.

Together, these experiments form a hybrid LLM and RL architecture in which strategy generation and execution are jointly modeled by individual agents.

Results & Discussion

4.1 | Experiment 1 Results

In the first experiment, we set out to address Objective 1 of this research: whether large language models (LLMs), when guided by structured prompting techniques, can generate trading strategies that are coherent, economically grounded, and aligned with pre-defined risk constraints.

This section presents the empirical results across four prompt versions (P 1 to P 4 in Table 4.1), as presented in Section 3.2.2 we are assessing their impact on quantitative metrics: SR, PPL, H_{LLM} , and Prompt 3 onward, qualitative: Expert Review Score (ERS). The Expert Reviewer surveys were introduced from Prompt 3 onward, as this was the point at which the prompt structure became sufficiently stable and interpretable to support HITL assessment. The composition of the reviewer cohort, including role, gender, and age range, is summarized in Appendix F (Table F.1) to contextualize the feedback presented below.

By examining these metrics under both risk and no-risk profiles, we test the extent to which structured prompting affects the quality and safety of the generated strategies.

Table 4.1: Prompt Versions for Experiment 1

Prompt Version	Description
P1	Baseline Prompt - Only static technical indicators and price data. Used as the control prompt.
P2	Baseline + ICM - Adds prior strategy decisions and outcomes to allow the model to reflect on past performance. Feature importance likert (1–3) are introduced.
P3	ICM + Instruction Decomposition + CoT - Modular breakdown into six reasoning blocks, ICL exemplars, and Final Reflection.
P4	Prompt 3 + News Factors - Adds textual macro and firm-specific news.

For the detailed survey responses, see Appendix F.

Table 4.2: Sharpe Ratio (SR) per Prompt and Ticker under Risk and No-Risk Profiles.

Ticker	Risk				No-Risk				Benchmark
	P1	P2	P3	P4	P1	P2	P3	P4	
AAPL	1.094	1.067	1.067	2.094	1.264	1.067	1.067	0.765	1.27
AMZN	0.348	0.380	0.631	0.844	-0.789	0.631	0.844	0.844	0.208
GOOGL	0.257	0.524	0.524	1.118	-0.278	0.524	0.524	0.524	0.19
META	-0.061	-0.276	0.296	0.766	0.041	-0.276	0.296	0.296	0.63
MSFT	1.069	1.108	1.306	0.498	0.431	1.306	1.306	1.770	1.17
TSLA	0.713	0.753	0.431	0.787	1.153	1.153	1.507	1.056	0.67
Mean	0.570	0.593	0.709	1.018	0.304	0.734	0.924	0.876	0.690

Table 4.3: Perplexity (PPL) per Prompt and Ticker for Risk Profiles.

Ticker	Risk				No-Risk			
	P1	P2	P3	P4	P1	P2	P3	P4
AAPL	1.851	1.313	1.550	1.444	1.684	1.356	1.469	1.343
AMZN	1.743	1.350	1.678	1.309	1.742	1.424	1.676	1.490
GOOGL	1.774	1.494	1.778	1.326	1.659	1.541	1.401	1.305
META	1.730	1.313	1.392	1.380	1.849	1.419	1.425	1.407
MSFT	1.833	1.439	1.494	1.237	1.770	1.305	1.482	1.246
TSLA	1.773	1.501	1.626	1.386	1.740	1.485	1.686	1.425
Mean	1.784	1.4017	1.5863	1.3470	1.7407	1.4217	1.5232	1.3693

Table 4.4: Entropy (H_{LLM}) per Prompt and Ticker for Risk Profiles.

Ticker	Risk				No-Risk			
	P1	P2	P3	P4	P1	P2	P3	P4
AAPL	0.7022	0.6740	0.6627	0.6851	0.6453	0.6630	0.7316	0.6324
AMZN	0.6934	0.6913	0.6928	0.6689	0.6493	0.6550	0.7033	0.6380
GOOGL	0.6678	0.6694	0.6956	0.6613	0.6621	0.6426	0.6642	0.6493
META	0.6640	0.6967	0.7314	0.6749	0.6418	0.6655	0.6876	0.6399
MSFT	0.6618	0.6797	0.7174	0.6487	0.6769	0.6618	0.7009	0.6449
TSLA	0.6751	0.6976	0.7414	0.6530	0.6234	0.6815	0.7041	0.6719
Mean	0.6774	0.6848	0.7069	0.6653	0.6498	0.6616	0.6986	0.6461

4.1.1 | Discussion

The results presented in Tables 4.2–4.6 provide empirical support for the effectiveness of structured prompt engineering in enhancing both the performance and certainty of LLM-generated trading strategies. Across the evaluated prompts, we observed consistent improvements in SRs, reductions in PPL and uncertainty (H_{LLM}), particularly as prompts progressed from unstructured templates to those incorporating decomposed reasoning and unstructured signals such as news.

Prompt 1 configuration yielded the weakest overall performance, with mean SR of 0.55, alongside moderate PPL and H_{LLM} values. We noted how Prompt 1, with the

Table 4.5: t -test p -values for Metric Changes Between P1 and P4 under Risk and No-Risk Profiles.

Metric	Risk	No-Risk
Sharpe Ratio	2.3×10^{-5}	1.6×10^{-1}
Perplexity	7.3×10^{-2}	9.4×10^{-2}
Entropy	2.3×10^{-4}	8.7×10^{-2}

Table 4.6: Expert Reviewer Scores (ERS) for the final Prompt 4 (1–3 rubric).

Evaluation Dimension	ERS
Rationale	2.7
Fidelity	2.65
Safety	2.8

tuned features and instructions led to decrease in the SR and elevated uncertainty, until further significant improvements were done in subsequent prompts to help the LLM rationalize its context, verified by the p -values in Table 4.5.

The only notable exception in tuning Prompt 1 was AAPL, discussed later in this section. Overall, the performance of the baseline and Prompt 1 was consistently poor across assets, underscoring the limited effectiveness of LLMs when unguided by structured prompting.

Prompt 2 introduced ICM in the form of past performance summaries. This led to performance gains, especially in the No-Risk setting, where the mean SR rose to 0.74. Several assets, including AMZN, GOOGL, and META, improved significantly. Modest reductions in PPL and entropy suggest that reflective priors helped the LLM achieve more confident completions. However, the verbosity and the lack of structured guidance to the LLM’s reasoning likely sustained elevated uncertainty seen in H_{LLM} .

Prompt 3 incorporated CoT reasoning with explicit decomposition over five domains (Stock, Technical, Macro, Options, News). This version achieved the highest SR in the No-Risk condition (0.92), and a competitive mean of 0.71 in the Risk setting. TSLA and MSFT notably benefited under risk, while AAPL and TSLA led under No-Risk. PPL increased slightly relative to Prompt 2, suggesting that deeper semantic scaffolds introduced more cognitive load.

Prompt 4 extended Prompt 3 by incorporating unstructured news. It achieved the highest mean SR under Risk (1.02), and the second highest under No-Risk (0.88) and exhibited the lowest average PPL and H_{LLM} values. The improved SR performance for

the high-risk profile supports the hypothesis that LLMs are well-suited to synthesizing textual sentiment, consistent with prior work on financial text modeling (Lopez-Lira & Tang, 2023; H. Yang et al., 2023; Zhou et al., 2021) though these same papers only performed short-term trading measured in days. The experienced Expert Reviewers' own trade rationales also prioritized market news in their reasoning, reinforcing these results.

At the individual asset's level, notably TSLA, prompt 4 SR outperform all prior prompts under Risk (0.79). This suggests that TSLA's sentiment-driven price dynamics benefited from narrative-sensitive prompting. Conversely, AAPL showed strong performance across all prompts except prompt 4 in a low-risk setting. Given the baseline's own SR performance (1.13), this indicates that for stable, trend-following and low-volatility assets, even minimal prompts can yield high-quality strategies.

Interestingly, Prompt 3 was most effective under No-Risk for most assets. This pattern suggests that the introduction of news-based inputs may inject detrimental short-term volatility for conservative profiles. This can be seen in comments from the investment focused Expert Reviewers, to paraphrase one: *"The company's strong fundamentals and market position outweigh the negative news, which is viewed as less significant given clear indicators of financial strength"*, an investor can choose to ignore the news and other short-term market-noise in favor of strong fundamentals. Thus, for stable strategy design with lower variance objectives, simpler prompts such as Prompt 1 may be preferred, this validated in Table 4.5, where the changes for No-Risk were not statistically significant in all three metrics, unlike the Risk profile where only PPL was not significant. Given the overall lack of significance for the PPL metric, the prompt quality and fluency has marginal benefits for the LLM, however, they offered substantial advantages for the expert reviewers in interpreting the model's decisions (see Section 3.2.2.9 for a discussion on why these metrics serve as proxies for prompt quality).

The discussed prompt metrics are the result of the prompt-mining done in Section 3.2.2.3. This method ensured that prompts captured relevant decision patterns that aligned with the exemplar trades provided. The most frequently cited features and explanatory bigrams used in the generated strategies under both high-risk and low-risk personas are summarized in Figures 4.1–4.4.

To better understand how the LLM internalizes financial reasoning under risk configurations, we analyzed the most frequently cited features and explanatory bigrams generated in response to high-risk prompts. Figures 4.1 and 4.2 report the top quartile of each distribution, providing a more interpretable subset of the most salient strategic signals.

Feature usage reveals a heavy emphasis on volatility-sensitive indicators such as

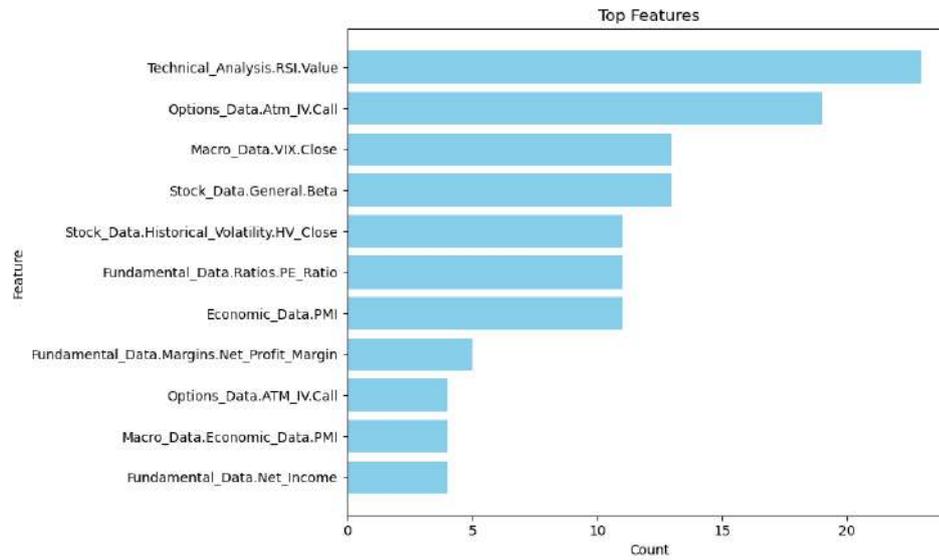


Figure 4.1: Top quartile features used under high-risk configuration.

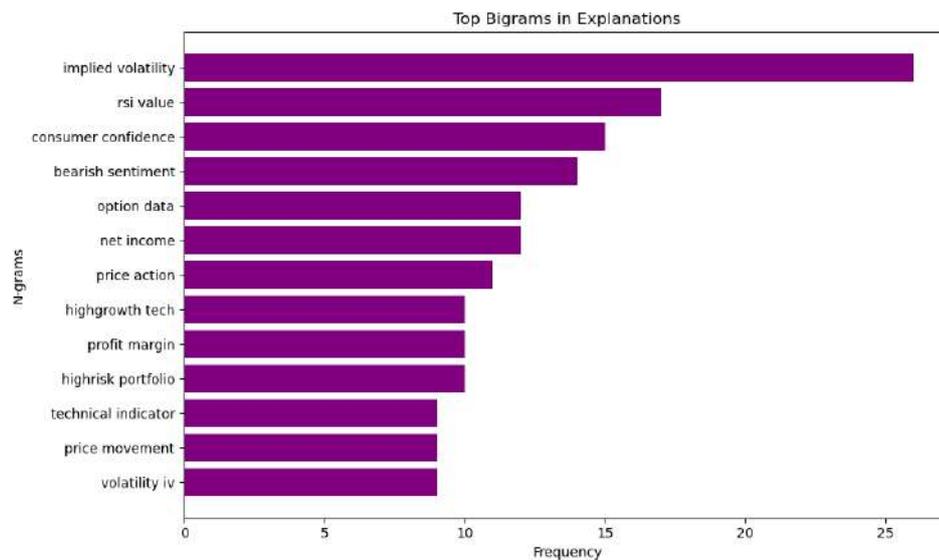


Figure 4.2: Top quartile bigrams in explanations under high-risk configuration

RSI, *Atm_IV.Call*, and *Beta*, as well as macroeconomic signals like *PMI* and *VIX*. This pattern suggests that the model preferred momentum-based, sensitivity-driven, and macro-surprise variables when constructing high-return, high-volatility strategies.

Bigrams in the accompanying rationales reflect this framing. Frequent expressions include *“implied volatility”*, *“decision take”*, *“take position”*, and *“option data”*, illustrating

a discourse centered on action-oriented, volatility-informed reasoning. Notably, terms such as *“high-growth tech”* and *“high-risk portfolio”* also appeared, indicating thematic alignment with speculative positioning. The coherence between feature salience and narrative elements implies that the model not only selected volatility-relevant inputs but also articulated them using language patterns consistent with tactical risk-seeking behavior.

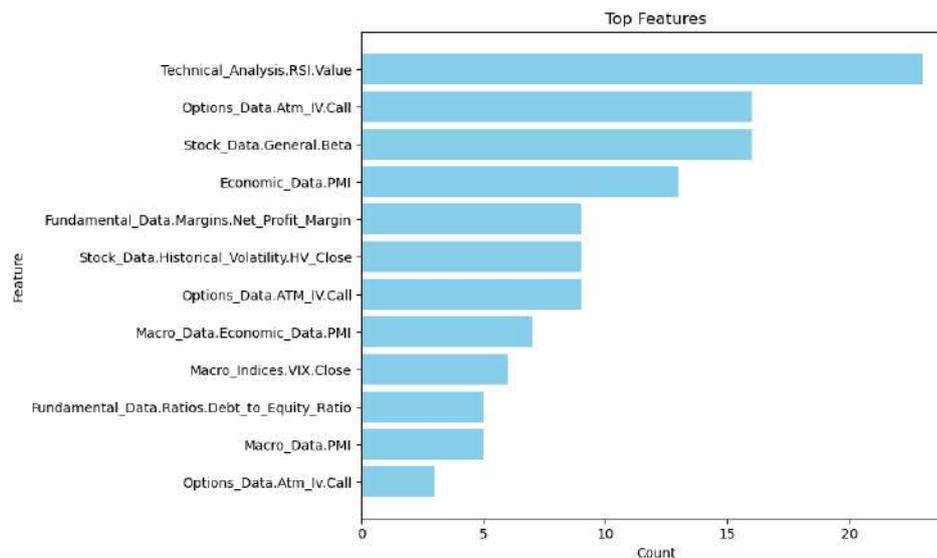


Figure 4.3: Top quartile features used under low-risk configuration

To examine how the LLM adapts its reasoning under low-risk configurations, we analyzed the most cited features and explanatory phrases from the low-risk strategy corpus. Figures 4.3 and 4.4 summarize the top quartile of features and bigrams, respectively, and emphasize a distinct risk-averse reasoning pattern.

While RSI again ranked highest, its prominence was balanced by greater emphasis on fundamental and macroeconomic indicators such as `Net_Profit_Margin`, `PE_Ratio`, and `PMI`. These features suggest a conservative orientation, with focus placed on earnings quality, financial robustness, and macroeconomic stability.

The shift in narrative structure was similarly apparent. Frequent bigrams included *“controlled volatility”*, *“negative return”*, *“low-risk trading”*, and *“profit margin”*, all indicative of a bias toward downside protection and capital preservation. Terms such as *“technical indicator”* and *“consumer confidence”* bridged both risk settings, reflecting the model’s consistent use of short-term indicators and sentiment data.

This reliance on price-based signals in both configurations is likely attributable to pretraining biases, as the underlying model was primarily exposed to general-purpose

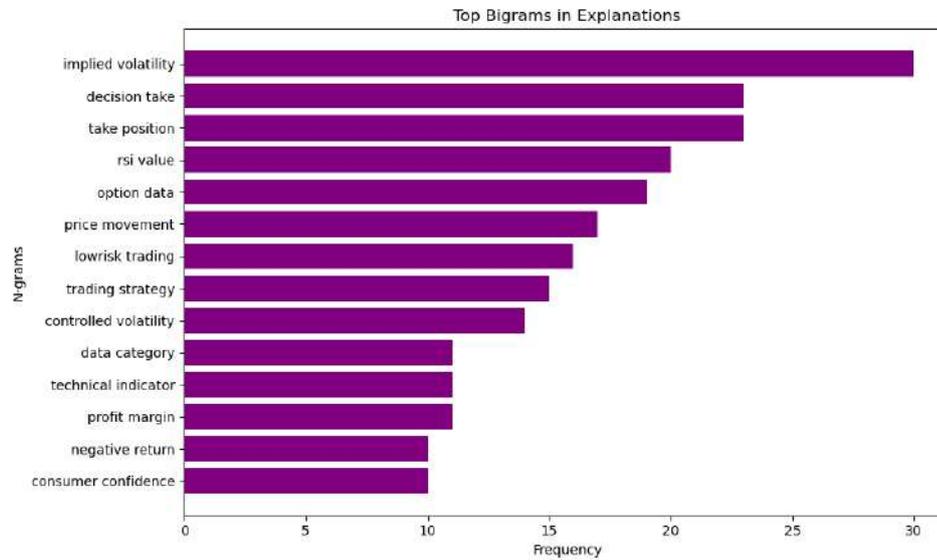


Figure 4.4: Top quartile bigrams in explanations under low-risk configuration

financial media rather than advanced financial theory. Consequently, while the model was capable of adapting its decision rationale to match risk personas when guided by structured prompts, its default feature preferences leaned toward technical heuristics commonly found in retail-focused trading discourse.

We also explored the potential of distillation, as the *Writer–Judge* loop produced a corpus of high-quality prompt–response pairs, comprising generated strategies and their corresponding evaluations by the Judge component. While this dataset could, in principle, support fine-tuning of a smaller language model, the associated resource costs—exceeding \$300 per training episode on OpenAI infrastructure—rendered this approach infeasible within the constraints of this study. Instead, we adopted an ICL paradigm as a practical alternative to explicit fine-tuning. As argued by Dai et al. (2023b), ICL enables large language models to achieve implicit task adaptation by conditioning on structured demonstrations embedded directly within the prompt. This approach exploits the model’s self-attention mechanism to generalize from few-shot exemplars without requiring parameter updates, making it a cost-effective and scalable solution for incorporating domain-aligned reasoning patterns. Therefore, we selected the highest-performing completions from earlier writer–judge iterations and reused them as ICL examples in our prompt generations along with exemplars from the expert reviewers.

Finally, the Expert Reviewer surveys provided critical insights that guided refine-

ments to Prompt 4. As shown in Table 4.6, reviewers rated the LLM's rationale as moderate (2.7 out of 3), praising its structured financial logic and ability to integrate sentiment, valuation, and regulatory inputs. One expert observed: *"Given the regulatory scrutiny, supply chain issues, and legal disputes creating short-term risks, I would have chosen shorting the stock... This is teaching me to also trust technical indicator analysis."* This kind of feedback affirmed the model's ability to synthesize structured and unstructured signals across domains.

Fidelity received a slightly lower average score (2.65), with reviewers noting occasional imprecision in the use of technical indicators. A representative critique stated: *"Calling RSI near 40 'approaching oversold' is debatable; oversold is typically <30."* Such feedback pointed to the need for more rigorous and domain-specific language in quantifying signals. Prompt 4 was revised accordingly to clarify technical thresholds and improve numerical accuracy.

Safety was rated the highest (2.8), with most reviewers acknowledging the LLM's balanced tone. However, a minority raised concerns about insufficient handling of complex regulatory and governance scenarios. One respondent noted: *"Model fails to capture regulatory, legal matters and company's bottleneck in this scenario."* These concerns informed further prompt tuning to extract and focus more on legal and compliance factors in the unstructured data.

Evaluations also varied by professional background. Reviewers from asset management often emphasized model transparency and signal prioritization. For example: *"The feature weighting is not very transparent... Is the LLM weighting price and technical analysis more than fundamental indicators?"* In contrast, both professional and retail traders tended to focus more on the alignment of technical and sentiment cues, as illustrated by: *"The bullish momentum of 20MA and 50MA with RSI indicating further upside potential. Yet the news on the company is concerning."*

In one particular instance, errors in the evaluation pipeline caused by temporarily missing input data, exposed the LLM's vulnerability to generating confabulations. A reviewer flagged several inconsistencies: *"(i) the model described a beta of 2.88 as indicative of aggressive trading suitable for a high-risk profile, while simultaneously characterizing the stock as defensive; (ii) it cited two conflicting RSI values (61.31 and 39.67) without reconciliation or explanation"*. This case highlighted the necessity of robust prompt design, and internal consistency checks as safeguards against misleading outputs. It also reinforced the value of incorporating intrinsic model evaluation to detect and mitigate confabulation risks.

Interestingly, the trust expressed by individuals in senior industry positions (3 of the 10 reviewers) was comparatively muted, centering around average scores. These re-

viewers frequently emphasized the need for additional data sources to support decision-making. Given their depth of experience, many were able to identify anonymized companies, which may have introduced bias-particularly toward one or two well-known firms.

The feedback led to several concrete modifications in Prompt 4, which were incorporated via ICL:

- Reinforced causal connections between narratives and signals (e.g., litigation, regulation, supply shocks).
- Calibrated language for expressing uncertainty and strategy confidence levels, e.g., "Long, with a confidence of 2 out of 3".

These adjustments ensured that Prompt 4 satisfied the three key criteria for real-world application. Taken together, the results validate the central claim of Experiment 1: when guided by structured prompts, LLMs can function as high-level strategic planners that generate coherent, expert-aligned financial reasoning.

Conclusions from Reviewers:

- Several experts emphasized the need for clearer numeric thresholds in technical analysis, particularly around RSI, MACD, and moving average crossovers.
- Signal weighting emerged as a key concern, especially for reviewers with buy-side experience who sought greater transparency around how fundamentals, price action, and sentiment are balanced.
- Some reviewers noted insufficient interpretive depth regarding news events, suggesting that richer sentiment modeling-especially around consumer behavior and macroeconomic shifts-could improve relevance.
- Several respondents proposed adding comparative benchmarking to peer companies and industry baselines as a way to contextualize model conclusions.
- Temporal ambiguity in trade rationale was flagged as a limitation. Future prompt structures should explicitly differentiate short-term tactical signals from longer-term strategic positioning.
- Finally, a recurring theme was that strong fundamentals, when paired with product-level momentum (e.g., user growth), often outweighed negative headlines. This suggests that prompts should allow models to prioritize structurally meaningful features when present.

Given its consistent outperformance across the test bench (Table 4.2), Prompt 4 will be used exclusively to generate the strategic policy envelope for the RL agent in Experiment 2.

4.2 | Experiment 2a Results - High Risk Profile

Experiment 2a examines whether LLM guidance improves performance in RL agents trading on a high-risk mandate, measured by the agent’s SR.

Table 4.7 presents the results for six individual equities, comparing the hybrid LLM+RL agent against two baselines: (i) a pure RL agent from the benchmark Théate and Ernst (2021), and (ii) the LLM-only SR from the high-risk Prompt 4 in Experiment 1.

Standard deviations are shown in parentheses. As outlined in Section 3.2.3.4, all RL metrics were evaluated over repeated runs with the same environment. To assess whether performance differences between the new architecture and the replicated baseline (see Appendix D.1) are statistically significant, we performed two-sided paired t -tests over 25 matched runs for each configuration.

The null hypothesis H_0 states that there is no difference in mean SR between the architectures: $H_0 : \mu_{\text{LLM+RL}} = \mu_{\text{RL-Only}}$. The alternative hypothesis H_1 states that a difference exists: $H_1 : \mu_{\text{LLM+RL}} \neq \mu_{\text{RL-Only}}$.

All reported p -values fall below 0.05 and indicate statistically significant differences under the 5% threshold.

Table 4.7: Experiment 2a. Sharpe Ratio (SR) for High-Risk Profile.

Note. Paired t -tests compare LLM+RL against RL-Only.

Ticker	LLM+RL	RL-Only	LLM-Only
TSLA	0.92 (0.22)	0.62 (0.60)	0.87
META	0.46 (0.70)	0.15 (0.61)	0.77
AMZN	1.21 (0.09)	0.42 (0.23)	0.84
GOOGL	1.16 (0.14)	0.23 (0.37)	1.12
MSFT	1.01 (0.21)	0.99 (0.30)	0.50
AAPL	1.703 (0.04)	1.42 (0.05)	2.09
Mean	1.077	0.639	1.032

4.2.1 | Discussion: Experiment 2a (High Risk)

Experiment 2a assessed whether LLM-guided RL agents could outperform traditional RL baselines under high-risk mandates, where maximizing returns is prioritized despite elevated volatility. Table 4.7 presents the results.

In five out of six assets, the LLM+RL hybrid agent achieved a higher SR than the RL benchmark. The largest absolute improvements were observed in AMZN and GOOGL, where the SR increased from 0.42 to 1.21 and from 0.23 to 1.16, respectively. These gains—representing nearly threefold improvements—indicate that the LLM contributed meaningful strategic guidance beyond what was discoverable by the RL agent alone. Notably, in both cases, the hybrid model also outperformed the LLM-only baseline, which achieved SRs of 0.84 (AMZN) and 1.12 (GOOGL), respectively.

For TSLA, the SR rose from 0.62 to 0.92 under the LLM+RL policy, closely aligning with the LLM-only benchmark of 0.87. This suggests that the hybrid agent effectively integrated the synthesized strategy, likely improving its capacity to navigate TSLA's frequent regime shifts and sentiment-sensitive dynamics.

META showed moderate gains, with the hybrid agent improving from 0.15 to 0.46. However, the LLM-only baseline outperformed both (0.77), suggesting that the RL component may not have fully leveraged the synthesized strategy, or was hindered by the short-term volatility in its reward function, suggesting a better reward function.

MSFT yielded a narrow margin between the RL baseline (0.99) and the LLM+RL agent (1.01), with the LLM-only model substantially underperforming (0.50). This suggests that the RL component played a dominant role in policy formation, and LLM guidance served as a minor influence.

AAPL exhibited a contrasting pattern. The LLM+RL agent outperformed the RL benchmark (1.70 vs. 1.42) but underperformed the LLM-only baseline (2.09). This warranted a deeper analysis, as it revealed a limitation in the RL's architecture. Figure 4.5 illustrates AAPL's trading behavior during one episode. The top panel plots price, technical indicators, and trades: hollow triangles mark RL trades; filled arrows show LLM monthly guidance. The LLM issued sparse but confident signals (strength > 0.6), often aligned with technical points of interest. In contrast, the RL agent frequently mistimed entries and exits.

From December 2018 to January 2019, the RL agent oscillated between LONG and SHORT positions with punishing results and despite receiving strong signals from the LLM. The LLM issued high-confidence guidance for a SHORT in December followed by a LONG in January, both with signal strengths exceeding 0.8. Regardless, the RL agent held a LONG position throughout the decline. As shown in Figure 4.7 (zoomed to show the values at the end of the training sessions), the DDQN assigns lower Q-values to SHORT actions (closer to 0) when compared to the LONG Q-value (closer to 0.005) in Figure 4.6, indicating limited confidence. This stems from the RL architecture in Théate and Ernst (2021) where the environment has lower-bound constraints (used to cap leverage from borrowing shares to SHORT) which created an asymmetric return



Figure 4.5: AAPL Performance with LLM+RL Model.

function by triggering buy-to-cover after price increases, reducing portfolio value and subsequent SHORT exposure. Also, the selected equity universe has positive historical drift, which constantly raises average prices and therefore exposing the RL to limited opportunities to learn the value of SHORT actions. Regardless of the RL's limitations, the LLM had a visible, albeit minor, performance increase as seen in the SR values in Figure 4.8 during the training sessions.

Taken together, these results support the central hypothesis of Experiment 2a: LLM-guided RL agents can enhance performance under high-risk configurations. The hybrid architecture proved particularly effective in volatile assets where long-term planning and sentiment interpretation were essential.

4.3 | Experiment 2b Results (Low Risk)

Experiment 2b evaluates the same architecture under a low-risk mandate, where the primary objective is to minimize downside exposure, measured by maximum draw-down (MDD). Unlike the return-maximizing configuration in Experiment 2a, this setup prioritizes risk-management and safety.

Table 4.8 presents MDD results across six equities, comparing the hybrid LLM+RL

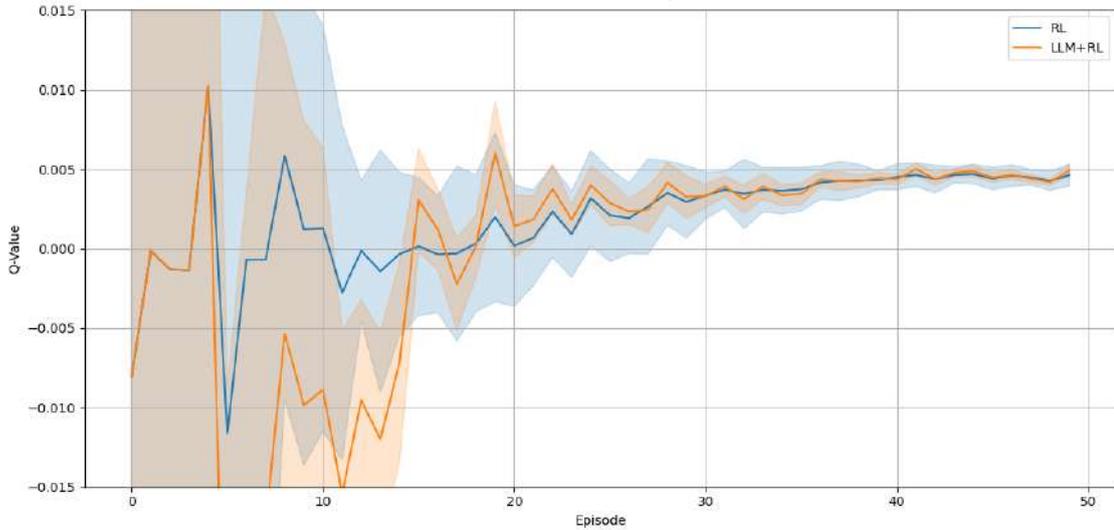


Figure 4.6: Training Behavior for AAPL: Q-Values for LONG.

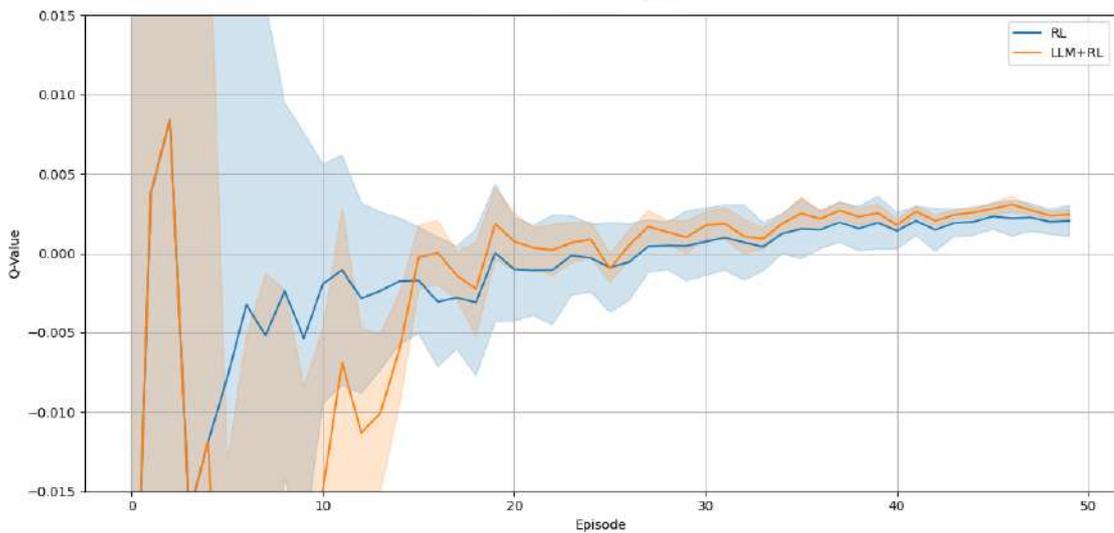


Figure 4.7: Training Behavior for AAPL: Q-Values for SHORT.

agent to both the RL-only baseline from Théate and Ernst (2021) and the best-performing low-risk LLM-only strategies (Prompt 4 from Experiment 1).

Standard deviations are reported in parentheses. All results were averaged over 25 runs using fixed seeds and environments. To assess the statistical significance of observed performance differences, we conducted two-sided paired t -tests comparing the drawdown metrics between agents across matched runs.

The null hypothesis H_0 states that the hybrid architecture has the same mean MDD

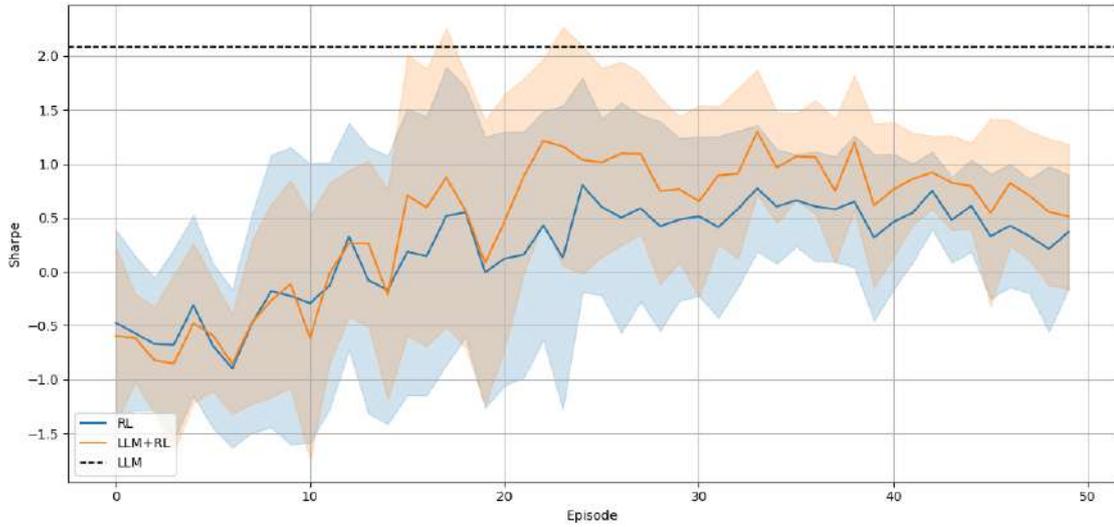


Figure 4.8: Training Behavior for AAPL: Sharpe Ratio.

as the replicated baseline (See Appendix D.1): $H_0 : \mu_{\text{LLM+RL}} = \mu_{\text{RL-Only}}$. The alternative hypothesis H_1 states that the mean MDD differs: $H_1 : \mu_{\text{LLM+RL}} \neq \mu_{\text{RL-Only}}$.

All reported p -values fall below 0.05 unless otherwise stated, indicating statistically significant differences under the 5% threshold.

Table 4.8: Experiment 2b. Maximum Drawdown (MDD) for Low-Risk Profile.

Ticker	LLM+RL	RL-Only	LLM-Only
TSLA	0.35 (0.02)	0.65 (0.13)	0.59
META	0.21 (0.21)	0.45 (0.27)	0.30
AMZN	0.32 (0.15)	0.19 (0.14)	0.34
GOOGL	0.19 (0.21)	0.25 (0.18)	0.35
MSFT	0.21 (0.08)	0.17 (0.09)	0.21
AAPL	0.22 (0.02)	0.45 (0.01)	0.28
Mean	0.250	0.360	0.345

4.3.1 | Discussion: Experiment 2b (Low Risk)

Experiment 2b tested the same agents under a low-risk mandate, with an emphasis on drawdown control via safe policies. Results are shown in Table 4.8, using MDD as the principal risk metric.

Across four of six assets, the LLM+RL agent achieved a lower MDD than the RL benchmark. The most substantial improvements were observed in TSLA and META, where MDD fell from 0.65 to 0.35 and from 0.45 to 0.21, respectively. This suggests

that the LLM conditioning effectively promoted loss-averse behavior, reducing exposure during high-volatility episodes and accelerating recovery-aligned with the safety-oriented design goal of the prompt.

GOOGL and AAPL exhibited modest drawdown reductions under the LLM+RL agent, though in both cases, the LLM-only (Prompt 4) strategy achieved lower or comparable MDDs. Notably, the hybrid architecture did not outperform the RL benchmark in AMZN or MSFT, where the local RL policy effectively preserved capital-evidenced by AMZN's benchmark MDD of 0.19 compared to 0.32 under the LLM+RL configuration. These asset-specific outcomes highlight the importance of reward shaping to adapt LLM-conditioned behavior to structurally diverse environments.

Averaged across assets, the LLM+RL policy consistently achieved better MDD than the RL benchmark, while remaining comparable to the LLM-only baseline. This reinforces the notion that in low-risk configurations, prompt content plays a dominant role in shaping agent behavior, with policy learning acting more as a stabilizer than a driver of new behavior. Crucially, this adaptive behavior is realized without modifying the underlying RL architecture; instead, the LLM dynamically adjusts strategic guidance to match low-risk investor preferences.

4.4 | Overall Findings

These experiments demonstrate that prompt-engineered strategies not only rival the performance of RL-only agents but also yield interpretable risk-aware policies, as validated by Expert Reviewers.

When deployed in a hybrid LLM+RL architecture, the system outperformed both the RL and LLM-only baselines in four out of six assets under both risk configurations.

This suggests a complementary relationship: the LLM provides strategic foresight by compensating for the RL agent's limited observation window and domain knowledge, while the RL contributes tactical precision in executing trades and short-term adaptability. Notably, the same underlying RL agent was used in both high- and low-risk experiments, with risk sensitivity emerging solely from the LLM's adaptation.

Conclusion

5.1 | Revisiting Aims and Objectives

This research set out to investigate whether LLMs can generate coherent, risk-aligned trading strategies, and whether those strategies can effectively guide RL agents toward safer or more performant behavior. These goals were defined through two research objectives and their corresponding experiments.

- **Objective 1: LLM Trading Strategy Generation.** The first objective assessed the capability of an LLM to produce trading strategies that are financially grounded, and aligned with specified risk personas. As demonstrated in Experiment 1 (Section 4.1), the LLM generated structured outputs that incorporated both quantitative indicators and unstructured signals. Among the prompt variants tested, Prompt 4 achieved the best overall performance, delivering consistently lower perplexity and entropy values with higher SR across assets and risk profiles. Expert Reviewers acting as HITL affirmed the internal consistency and strategic relevance of the generated explanations, further validating the LLM's ability to emulate domain-appropriate reasoning.
- **Objective 2: LLM-Guided RL Agent.** The second objective explored whether strategies derived from the LLM could improve the learning and behavior of downstream RL agents. As seen in Experiment 2a (Section 4.2), under a high-risk configuration, the LLM-guided agent outperformed the RL benchmark in four out of six assets. In Experiment 2b (Section 4.3), under a low-risk configuration, the hybrid agent demonstrated improved capital preservation with lower MDD in also four out of six assets. Importantly, both risk profiles were handled using the same underlying RL agent architecture, with the LLM adapting its strategies to meet the distinct risk constraints.

The results confirm the central hypothesis of this thesis: that LLMs, through well engineered prompts, can operate as an effective strategist, and can guide RL agents to produce risk-aware trading behavior that can yield a good trading performance.

5.2 | Critique and Limitations

While this research demonstrated promising outcomes in aligning LLM-generated strategies with RL agents, several limitations merit consideration.

First, the experimental framework was evaluated on a limited set of equities within the technology sector only, restricting the generalization of findings to other asset classes, sectors, and geographies under broader market regimes. The risk personas were also not explicitly defined; instead, the LLM inferred risk preferences based on general financial knowledge from its pretraining. No fine-tuning was performed; the LLM was guided solely through ICL examples, a choice dictated only by resource constraints.

Second, resource constraints significantly limited the ability to iterate and modulate experimental configurations. Running the full pipeline from LLM strategy generation to RL evaluation required approximately one day per prompt version and risk persona across the testbench, imposing a bottleneck on exploration and tuning. Each run costed approximately \$3-\$5 on OpenAI GPT-4o mini.

Third, the agentic system was simplified for feasibility. In production settings, decision-making in quantitative trading incorporates intractable data sources, such as satellite imagery, bond market rate dynamics, Limit Order Book (LOB) microstructure, local and global liquidity, asset correlations, fine-grain economic data, and extensive derivatives data. Modeling such complexity with only three LLM agents is inadequate. Furthermore, institutional systems typically optimize over portfolios, whereas this study focused on single-asset strategies to enable fair comparison with the benchmark study Théate and Ernst (2021). The choice of monthly strategy generation also constrained the LLM's responsiveness to data pivots, a limitation mainly driven by resource constraints.

Finally, this research primarily centered on the design and integration of LLM-based agents, with the RL component serving only as a tactical agent to be guided. Enhancements to the RL layer were beyond scope but represent a direction for future work. For instance, Takara et al. (2023) used exponential liquidation-penalty rewards, which better reflect realistic trading behavior than unrealized return signals. Advanced actor-critic methods such as SAC (Liu, Yang, et al., 2022), PPO, and hybrid architectures (Pricope, 2021; Xu et al., 2023) offer improved robustness and SR. Integrating these would likely improve tactical execution performance, synergized by the LLM-guided setup.

5.3 | Future Work

The findings of this research open several promising directions for further research at the intersection of large language models and reinforcement learning in quantitative finance.

First, more involved reward shaping. Having the LLM to participate in the reward is an ambitious extension where future systems could allow the LLM to revise or critique the reward function in response to required strategic outcomes and the current RL performance. This would open a path toward adaptive, value-aligned reward engineering that doesn't only guide, but controls the RL learning process to align more closely with a professional trader behaviors. A step more than reward shaping would be choosing what trained-RL agent to activate, if to use one at all, treating the agent as an Option in a HRL architecture.

Second, the strategy generation pipeline can be enhanced by broadening the set of available market signals. While the current framework draws on price, volatility, and macroeconomic features, future implementations may incorporate many datasets mentioned in the Limitations section above. Even better, this should be followed with more specialized LLM agents to reason over different or cross-sectional datasets, improving its ability to contextualize decisions under uncertainty. For instance, separate agents could be tasked with generating entry signals, calibrating position sizing, performing macro overlay adjustments, or managing risk constraints. Such a modular architecture would enable each agent to focus on a distinct subdomain, further improving its safety and trading metrics.

Third, portfolio-level learning and risk management should be explored. While single-asset strategies enabled controlled comparisons in an academic context, in the domain, portfolio of assets of various correlations are traded in sophisticated strategies. Portfolio-aware prompts and multi-agent coordination mechanisms could allow LLMs to scale to run multi-asset strategies.

Fourth, future research should examine the temporal distance between the experimental dataset (or live environment) and the model's knowledge cutoff. It remains an open question when historical distance becomes detrimental to model outputs and how to determine the appropriate point for retraining or fine-tuning.

5.4 | Final Remarks

This research has contributed to the emerging paradigm of LLM-guided RL in algorithmic trading by developing and validating a hybrid architecture in which LLMs serve as strategic planners and RL agents function as tactical executors. By leveraging structured prompt engineering, the system elicits coherent, risk-aware, and economically grounded trading strategies that are both interpretable and adaptable to real-world conditions.

Through the experiment performed, the research demonstrated that LLM-generated strategies, when correctly prompted, can significantly improve the performance and safety of downstream RL agents. Experiment 1 established that iterative prompt engineering alone can shape the quality and safety of LLM outputs, while Experiment 2 showed that these strategies, once deployed into the RL agent's observation space, enhance performance across both high- and low-risk configurations. Importantly, these results were achieved using the same underlying RL architecture, with all adaptations to risk profiles handled solely through the LLM-strategic guidance.

The findings support a paradigm shift away from monolithic RL architectures toward hybrid systems guided by domain- and risk- aligned LLMs, capable of emulating expert-level reasoning under varying market environments.

All source code developed for this study is publicly available on GitHub⁸, and instructions on running the code are available in the repository's README. While most datasets are accessible through public endpoints and can be retrieved using the provided scraping scripts, the data itself is not shared directly to ensure compliance with fair use policies and vendor usage agreements.

⁸<https://github.com/adamd1985/LLM-Guided-RL-in-Quant-Trading>

References

- Alonso, M., & Dupouy, H. (2024). Evaluating LLMs in financial tasks – Code generation in trading strategies. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4752797>
- Alsini, R., Al-Haija, Q. A., Alsulami, A. A., Alturki, B., Alqurashi, A. A., Mashat, M. D., Alqahtani, A., & Alhebaishi, N. (2024). Forecasting cryptocurrency's buy signal with a bagged tree learning approach to enhance purchase decisions. *Frontiers in Big Data*, 7, 1369895. <https://doi.org/10.3389/fdata.2024.1369895>
- Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, 34(6), 26–38. <https://doi.org/10.1109/msp.2017.2743240>
- Bailey, D. H., & López de Prado, M. (2014). The deflated Sharpe ratio: Correcting for selection bias, backtest overfitting and non-normality. *Journal of Portfolio Management*, 40(5), 94–107. <https://doi.org/10.2139/ssrn.2460551>
- Bartram, S. M., Branke, J., & Motahari, M. (2020). *Artificial intelligence in asset management*. CFA Institute Research Foundation. <https://doi.org/10.2139/ssrn.3692805>
- Booth, S., Knox, W. B., Shah, J., Niekum, S., Stone, P., & Allievi, A. (2023). The perils of trial-and-error reward design: Misdemeanor through overfitting and invalid task specifications. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(5), 5920–5929. <https://doi.org/10.1609/aaai.v37i5.25733>
- Bygrave, L. A. (2020, February). Automated individual decision-making, including profiling. In *The EU general data protection regulation (GDPR): A commentary*. Oxford University Press. <https://doi.org/10.1093/oso/9780198826491.003.0055>

- Cao, Y., Zhao, H., Cheng, Y., Shu, T., Liu, G., Liang, G., Zhao, J., & Li, Y. (2024). Survey on large language model-enhanced reinforcement learning: Concept, taxonomy, and methods. <https://doi.org/10.1109/TNNLS.2024.3497992>
- Chaddha, A., & Yadav, S. (2022). Examining the predictive power of moving averages in the stock market. *Journal of Student Research*, 11(3). <https://doi.org/10.47611/jsrhs.v11i3.3382>
- Chan, E. (2013). *Algorithmic trading: Winning strategies and their rationale*. John Wiley & Sons.
- Chlistalla, M. (2011, February). *High-frequency trading and long-term investment: A systematic review* (Research Briefing). Deutsche Bank Research. Frankfurt am Main, Germany. <https://doi.org/10.2139/ssrn.2066884>
- Commission, E. (2021). Laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts.
- Corazza, M., & Bertoluzzo, F. (2014, January). *Q-Learning-based financial trading systems with applications* (Working Papers No. 2014:15). Department of Economics, University of Venice "Ca' Foscari". <https://doi.org/dx.doi.org/10.2139/ssrn.2507826>
- Dai, D., Sun, Y., Dong, L., Hao, Y., Ma, S., Sui, Z., & Wei, F. (2023a). Why Can GPT learn in-context? Language models Implicitly Perform Gradient Descent as meta-optimizers. <https://doi.org/10.48550/arXiv.2212.10559>
- Dai, D., Sun, Y., Dong, L., Hao, Y., Ma, S., Sui, Z., & Wei, F. (2023b). Why Can GPT learn in-context? Language models secretly perform gradient descent as meta-optimizers. *Findings of the Association for Computational Linguistics: ACL 2023*, 4005–4019. <https://doi.org/10.18653/v1/2023.findings-acl.247>
- Dakalbab, F., Abu Talib, M., Nasir, Q., & Saroufil, T. (2024). Artificial intelligence techniques in financial trading: A systematic literature review. *Journal of King Saud University – Computer and Information Sciences*, 36, 102015. <https://doi.org/10.1016/j.jksuci.2024.102015>

- Dalal, M., Chiruvolu, T., & Chaplot, D. (2024). Plan-seq-learn: Language model guided RL for solving long horizon robotics tasks. *arXiv preprint arXiv:2405.01534*. <https://doi.org/10.48550/arXiv.2405.01534>
- Dang, Q.-V. (2020). Reinforcement learning in stock trading. *Advanced Computational Methods for Knowledge Engineering*, 311–322. https://doi.org/10.1007/978-3-030-38364-0_28
- Demajo, L. M., Vella, V., & Dingli, A. (2020). Explainable AI for interpretable credit scoring. <https://doi.org/10.5121/csit.2020.101516>
- Devidze, R., Kamalaruban, P., & Singla, A. K. (2022). Exploration-guided reward shaping for reinforcement learning under sparse rewards. *Proceedings of the Conference on Neural Information Processing Systems*. <https://doi.org/10.5555/3600270.3600692>
- Ding, H., Li, Y., Wang, J., & Chen, H. (2024). Large language model agent in financial trading: A survey. *arXiv preprint arXiv:2408.06361*. <https://doi.org/10.48550/arXiv.2408.06361>
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383–417. <https://doi.org/10.1111/j.1540-6261.1970.tb00518.x>
- Fama, E. F., & French, K. R. (2004, September). *The capital asset pricing model: Theory and evidence* (Vol. 18). <https://doi.org/10.1257/0895330042162430>
- Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1–22. <https://doi.org/10.1016/j.jfineco.2014.10.010>
- Gonen, H., Iyer, S., Blevins, T., Smith, N., & Zettlemoyer, L. (2023). Demystifying prompts in language models via perplexity estimation. *Findings of the Association for Computational Linguistics: EMNLP 2023*, 10136–10148. <https://doi.org/10.18653/v1/2023.findings-emnlp.679>
- Han, C., He, Z., & Toh, A. J. W. (2023). Pairs trading via unsupervised learning. *European Journal of Operational Research*, 307(2), 929–947. <https://doi.org/10.1016/j.ejor.2022.09.041>

- Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2018). *Metrics for explainable AI: Challenges and prospects* (Technical Report). DARPA XAI Program. <https://doi.org/10.48550/arXiv.1812.04608>
- Hu, Y., Wang, X., Yao, W., Lu, Y., Zhang, D., Foroosh, H., Yu, D., & Liu, F. (2024). DeFine: Enhancing LLM Decision-Making with Factor Profiles and Analogical Reasoning. <https://doi.org/10.48550/arXiv.2410.01772>
- Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research*, 40(2), 806–841. <https://doi.org/10.1111/1911-3846.12795>
- Huang, J., & Chang, K. C.-C. (2023). Towards reasoning in large language models: A survey. <https://doi.org/10.48550/arXiv.2212.10403>
- Joubert, J., Sestovic, D., Barziy, I., Distaso, W., & López de Prado, M. (2024). The three types of backtests. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4897573>
- Kaltchenko, A. (2025). Entropy heat-mapping: Localizing GPT-based OCR errors with Sliding-Window Shannon analysis. *arXiv preprint arXiv:2505.00746*. <https://doi.org/10.48550/arXiv.2505.00746>
- Lazaridis, A., Fachantidis, A., & Vlahavas, I. (2020). Deep reinforcement learning: A state-of-the-art walkthrough. *Journal of Artificial Intelligence Research*, 69, 1421–1471. <https://doi.org/10.1613/jair.1.12197>
- Lei, B., & Song, Y. (2024). Volatility forecasting for stock market incorporating media reports, investors' sentiment, and attention based on MTGNN model. *Journal of Forecasting*, 43(3), 1706–1730. <https://doi.org/10.1002/for.3101>
- Li, Y. (2017). Deep reinforcement learning: An overview. *CoRR*, abs/1701.07274. <https://doi.org/10.48550/arXiv.1701.07274>
- Liu, X.-Y., Rui, J., Gao, J., Yang, L., Yang, H., Wang, Z., Wang, C. D., & Guo, J. (2022). FinRL-meta: A universe of near-real market environments for data-driven deep reinforcement learning in quantitative finance. <https://doi.org/10.48550/arXiv.2112.06753>

- Liu, X.-Y., Yang, H., Chen, Q., Zhang, R., Yang, L., Xiao, B., & Wang, C. D. (2022). FinRL: A deep reinforcement learning library for automated stock trading in quantitative finance. <https://doi.org/10.48550/arXiv.2011.09607>
- Liu, X.-Y., Yang, H., Gao, J., & Wang, C. D. (2021). FinRL: Deep reinforcement learning framework to automate trading in quantitative finance. *Proceedings of the Second ACM International Conference on AI in Finance*, 601–608. <https://doi.org/10.1145/3490354.3494366>
- López de Prado, M. (2018a). *Advances in financial machine learning*. Wiley.
- López de Prado, M. (2018b, January). *The 10 reasons most machine learning funds fail* (tech. rep.). True Positive Technologies. <https://doi.org/10.2139/ssrn.3104816>
- López de Prado, M. (2020a). Beyond econometrics: A roadmap towards financial machine learning. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3365282>
- López de Prado, M. (2020b). *Machine learning for asset managers*. Cambridge University Press.
- Lopez-Lira, A., & Tang, Y. (2023, May). Can ChatGPT forecast stock price movements? Return predictability and large language models. <https://doi.org/10.48550/arXiv.2304.07619>
- Lopez-Lira, A., Tang, Y., & Zhu, M. (2025). The memorization problem: Can We Trust LLMs' Economic Forecasts? <https://doi.org/10.48550/arXiv.2504.14765>
- Mai, D. (2024). StockGPT: A GenAI model for stock prediction and trading. *SSRN Electronic Journal*. <https://doi.org/10.48550/arXiv.2404.05101>
- Mezghani, L., Bojanowski, P., Alahari, K., & Sukhbaatar, S. (2023). Think before you act: Unified policy for interleaving language reasoning with actions. <https://doi.org/10.48550/arXiv.2304.11063>
- Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing Atari with Deep Reinforcement Learning. <https://doi.org/10.48550/arXiv.1312.5602>

- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*, 529–533. <https://doi.org/10.1038/nature14236>
- Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics*, *104*(2), 228–250. <https://doi.org/10.1016/j.jfineco.2011.11.003>
- Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2024). A comprehensive overview of large language models. <https://doi.org/10.48550/arXiv.2307.06435>
- Onozo, L., Arthur, F., & Gyires-Tóth, B. (2024). Leveraging LLMs for financial news analysis and macroeconomic indicator nowcasting. *IEEE Access*, *12*, 1–1. <https://doi.org/10.1109/access.2024.3488363>
- OpenAI. (2024). GPT-4o mini: Advancing cost-efficient intelligence [Accessed: 2024-11-14]. <https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/>
- Parliament, E., & of the European Union, C. (2014). Directive 2014/65/EU of the European Parliament and of the Council on Markets in Financial Instruments. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32014L0065>
- Pateria, S., Subagdja, B., Tan, A.-H., & Quek, C. (2021). Hierarchical reinforcement learning: A comprehensive survey. *ACM Computing Surveys*, *54*(5), 1–35. <https://doi.org/10.1145/3453160>
- Pricope, T.-V. (2021). Deep reinforcement learning in quantitative algorithmic trading: A review. <https://doi.org/10.48550/arXiv.2106.00123>
- Pternea, M., Singh, P., Chakraborty, A., Oruganti, Y., Milletari, M., Bapat, S., & Jiang, K. (2024). The RL/LLM taxonomy tree: Reviewing Synergies Between Reinforcement Learning and Large Language Models. *Journal of Artificial Intelligence Research*, *80*. <https://doi.org/10.1613/jair.1.15380>

- Sánchez Serrano, A. (2020). High-frequency trading and systemic risk: A structured review of findings and policies. *Review of Economics*, 71(3), 169–195. <https://doi.org/10.1515/roe-2020-0028>
- Schulhoff, S., Ilie, M., Balepur, N., Kahadze, K., Liu, A., Si, C., Li, Y., Gupta, A., Han, H., Schulhoff, S., et al. (2024). The Prompt Report: A Systematic Survey of Prompting Techniques. <https://doi.org/10.48550/arXiv.2406.06608>
- Sharpe, W. F. (1966). Mutual fund performance. *The Journal of Business*, 39(1), 119–138. <https://doi.org/10.1086/294846>
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd). MIT Press. <http://incompleteideas.net/book/the-book-2nd.html>
- Takara, L., Santos, A., Mariani, V., & Coelho, L. (2023). Deep reinforcement learning applied to a sparse-reward trading environment with intraday data. *Expert Systems with Applications*, 238, 121897. <https://doi.org/10.1016/j.eswa.2023.121897>
- Tang, X., Xu, S., & Ye, H. (2022). The way to invest: Trading Strategies Based on ARIMA and Investor Personality. *Symmetry*, 14, 2292. <https://doi.org/10.3390/sym14112292>
- Théate, T., & Ernst, D. (2021). An application of deep reinforcement learning to algorithmic trading. *Expert Systems with Applications*, 173, 114632. <https://doi.org/10.1016/j.eswa.2021.114632>
- Vakil, S., Shah, M., & Rawal, N. (2024). Preferences in algorithmic trading: Current Practices and Future Innovations. <https://doi.org/10.52783/eel.v14i3.1865>
- van Hasselt, H., Guez, A., & Silver, D. (2015). Deep reinforcement learning with Double Q-learning. *arXiv preprint arXiv:1509.06461*. <https://doi.org/10.48550/arXiv.1509.06461>
- Vanstone, B., & Hahn, T. (2015). First steps in developing high-frequency trading models. *The Journal of Trading*, 10(2), 54–71. <https://doi.org/10.3905/jot.2015.10.2.054>

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. *CoRR*, *abs/1706.03762*. <https://doi.org/10.48550/arXiv.1706.03762>
- Wang, M., Izumi, K., & Sakaji, H. (2024). LLMFactor: Extracting Profitable Factors through Prompts for Explainable Stock Movement Prediction. <https://doi.org/10.48550/arXiv.2406.10811>
- Wang, S., Yuan, H., Ni, L. M., & Guo, J. (2024). QuantAgent: Seeking Holy Grail in Trading by Self-Improving Large Language Model. <https://doi.org/10.48550/arXiv.2402.03755>
- Wang, X., et al. (2020). Improving generalization in Reinforcement Learning with Mixture Regularization. *Advances in Neural Information Processing Systems*.
- Wang, Y., & Yan, G. (2021). Survey on the application of deep learning in algorithmic trading. *Data Science in Finance and Economics*, *1(4)*, 345–361. <https://doi.org/10.3934/dsfe.2021019>
- Xie, T., Zhao, S., Wu, C. H., Liu, Y., Luo, Q., & Zhong, V. (2023). Text2Reward: Reward Shaping with Language Models for Reinforcement Learning. *arXiv preprint arXiv:2309.11489*. <https://doi.org/10.48550/arXiv.2309.11489>
- Xing, F. (2025). Designing heterogeneous LLM agents for financial sentiment analysis. *ACM Transactions on Management Information Systems*, *16(1)*, 5. <https://doi.org/10.1145/3688399>
- Xu, M., Lan, Z., Tao, Z., Du, J., & Ye, Z. (2023). Deep reinforcement learning for quantitative trading. <https://doi.org/10.48550/arXiv.2312.15730>
- Yang, H., Liu, X.-Y., & Wang, C. D. (2023). FinGPT: Open-source financial large language models. <https://doi.org/10.48550/arXiv.2306.06031>
- Yang, H., Zhang, B., Wang, N., Guo, C., Zhang, X., Lin, L., Wang, J., Zhou, T., Guan, M., Zhang, R., & Wang, C. D. (2024). FinRobot: An open-source AI agent platform for financial applications using large language models. <https://doi.org/10.48550/arXiv.2405.14767>

- Yang, Y., Uy, M. C. S., & Huang, A. (2020). FinBERT: A Pretrained Language Model for Financial Communications. <https://doi.org/10.48550/arXiv.2006.08097>
- Yona, G., Aharoni, R., & Geva, M. (2024). Can large language models faithfully express their intrinsic uncertainty in words? *arXiv preprint arXiv:2405.16908*. <https://doi.org/10.48550/arXiv.2405.16908>
- Yoon, J., & Fan, J. (2024). Forecasting the direction of the Fed's monetary policy decisions using random forest. *Journal of Forecasting*, 43(7), 2848–2859. <https://doi.org/10.1002/for.3144>
- Yu, Y., Li, H., Chen, Z., Jiang, Y., Li, Y., Zhang, D., Liu, R., Suchow, J. W., & Khashanah, K. (2024). FinMem: A Performance-Enhanced LLM Trading Agent with Layered Memory and Character Design. *AAAI Spring Symposia*, 595–597. <https://doi.org/10.48550/arXiv.2311.13743>
- Yuan, H., Zhang, C., Wang, H., Xie, F., Cai, P., Dong, H., & Lu, Z. (2023). Skill Reinforcement Learning and Planning for Open-World Long-Horizon Tasks. <https://doi.org/10.48550/arXiv.2303.16563>
- Zhang, W., Zhao, L., Xia, H., Sun, S., Sun, J., Qin, M., Li, X., Zhao, Y., Zhao, Y., Cai, X., Zheng, L., Wang, X., & An, B. (2024). A multimodal foundation agent for financial trading: Tool-Augmented, Diversified, and Generalist. <https://doi.org/10.48550/arXiv.2402.18485>
- Zhou, Z., Ma, L., & Liu, H. (2021). Trade the Event: Corporate Events Detection for News-Based Event-Driven Trading. *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 2114–2124. <https://doi.org/10.18653/v1/2021.findings-acl.186>

Algorithms

This appendix describes all algorithms used or studied.

Algorithm 1: Deep Q-Learning with Experience Replay**Data:** Replay memory D of capacity N **Result:** Learned Q-function Q with parameters θ Initialize action-value function Q with random weights θ ;Initialize target action-value function \hat{Q} with weights $\theta^- = \theta$;**for** $episode = 1$ to M **do** Initialize sequence $s_1 = \{x_1\}$ and preprocessed input $\phi_1 = \phi(s_1)$; **for** $t = 1$ to T **do** With probability ϵ , select a random action a_t ;

Otherwise select

$$a_t = \arg \max_a Q(\phi(s_t), a; \theta)$$

 Execute action a_t in the environment and observe reward r_t and image x_{t+1} ; Set $s_{t+1} = s_t, a_t, x_{t+1}$ and preprocess $\phi_{t+1} = \phi(s_{t+1})$; Store transition $(\phi_t, a_t, r_t, \phi_{t+1})$ in D ; Sample random minibatch of transitions $(\phi_j, a_j, r_j, \phi_{j+1})$ from D ;

Set

$$y_j = \begin{cases} r_j & \text{if episode terminates at step } j + 1 \\ r_j + \gamma \max_{a'} \hat{Q}(\phi_{j+1}, a'; \theta^-) & \text{otherwise} \end{cases} \quad (\text{A.1})$$

Perform a gradient descent step on the loss:

$$(y_j - Q(\phi_j, a_j; \theta))^2 \quad (\text{A.2})$$

 Every C steps, update target network:

$$\theta^- \leftarrow \theta$$

end**end**

Algorithm 2: Double DQN Algorithm

Input: \mathcal{D} – empty replay buffer; θ – initial network parameters; θ^- – target network copy;

N_r – replay buffer max size; N_b – training batch size; N – target update frequency

for each episode $e = 1, 2, \dots, M$ **do**

 Initialize frame sequence $\mathbf{x} \leftarrow ()$;

for $t = 0, 1, 2, \dots$ **do**

 Set state $s \leftarrow \mathbf{x}$;

 Select action $a \sim \pi_\epsilon$ (e.g., ϵ -greedy);

 Execute a , observe next frame x^t , reward r , append x^t to \mathbf{x} ;

if $|\mathbf{x}| > N_f$ **then**

 | Remove oldest frame $x_{t_{\min}}$ from \mathbf{x}

end

 Set $s' \leftarrow \mathbf{x}$ and store transition (s, a, r, s') in \mathcal{D} ;

if $|\mathcal{D}| > N_r$ **then**

 | Replace oldest tuple in \mathcal{D}

end

 Sample minibatch of N_b tuples $(s, a, r, s') \sim \text{Unif}(\mathcal{D})$;

for each sampled tuple **do**

 Define $a^{\max}(s') \leftarrow \arg \max_{a'} Q(s', a'; \theta)$;

 Compute target:

$$y_j = \begin{cases} r, & \text{if } s' \text{ is terminal} \\ r + \gamma Q(s', a^{\max}(s'); \theta^-), & \text{otherwise} \end{cases}$$

end

 Perform gradient descent on:

$$\mathcal{L}(\theta) = \|y_j - Q(s, a; \theta)\|^2$$

if $t \bmod N = 0$ **then**

 | Update target network: $\theta^- \leftarrow \theta$

end

end

end

Algorithm 3: Expert Trade Heuristic

Data: Time-indexed price series**Result:** Trade action: LONG (1) or SHORT (0)**foreach** date t in dataset **do** $P_t \leftarrow \text{Close}(t)$; $r^{(10)} \leftarrow \frac{P_{t+10}}{P_t} - 1, r^{(20)} \leftarrow \frac{P_{t+20}}{P_t} - 1$; $r^{\text{weighted}} \leftarrow 0.4 \cdot r^{(10)} + 0.6 \cdot r^{(20)}$; **if** $r^{\text{weighted}} \geq 0$ **then** | **Action** \leftarrow LONG (Trade_Action = 1); **else** | **Action** \leftarrow SHORT (Trade_Action = 0); **end****end**

Equations

This appendix lists all financial ratios and technical analysis calculated for the dataset used.

B.1 | Financial and Technical Formulas

Liquidity Ratios

Quick Ratio

$$\text{Quick Ratio} = \frac{\text{Current Assets} - \text{Inventory}}{\text{Current Liabilities}} \quad (\text{B.1})$$

Current Ratio

$$\text{Current Ratio} = \frac{\text{Current Assets}}{\text{Current Liabilities}} \quad (\text{B.2})$$

Leverage and Coverage Ratios

Debt-to-Equity

$$\text{Debt-to-Equity} = \frac{\text{Total Debt}}{\text{Shareholders' Equity}} \quad (\text{B.3})$$

Debt-to-Assets

$$\text{Debt-to-Assets} = \frac{\text{Total Debt}}{\text{Total Assets}} \quad (\text{B.4})$$

Interest Coverage

$$\text{Interest Coverage} = \frac{\text{EBIT}}{\text{Interest Expense}} \quad (\text{B.5})$$

Profitability Ratios

Gross Margin

$$\text{Gross Margin} = \frac{\text{Gross Profit}}{\text{Revenue}} \quad (\text{B.6})$$

Operating Margin

$$\text{Operating Margin} = \frac{\text{Operating Income}}{\text{Revenue}} \quad (\text{B.7})$$

Net Profit Margin

$$\text{Net Profit Margin} = \frac{\text{Net Income}}{\text{Revenue}} \quad (\text{B.8})$$

Return on Equity (ROE)

$$\text{ROE} = \frac{\text{Net Income}}{\text{Shareholders' Equity}} \quad (\text{B.9})$$

Return on Assets (ROA)

$$\text{ROA} = \frac{\text{Net Income}}{\text{Total Assets}} \quad (\text{B.10})$$

Efficiency Ratios**Inventory Turnover**

$$\text{Inventory Turnover} = \frac{\text{COGS}}{\text{Average Inventory}} \quad (\text{B.11})$$

Asset Turnover

$$\text{Asset Turnover} = \frac{\text{Revenue}}{\text{Total Assets}} \quad (\text{B.12})$$

Fixed Asset Turnover

$$\text{Fixed Asset Turnover} = \frac{\text{Revenue}}{\text{Net Fixed Assets}} \quad (\text{B.13})$$

Valuation Ratios**Price-to-Earnings (P/E)**

$$\text{PE} = \frac{\text{Stock Price}}{\text{EPS}} \quad (\text{B.14})$$

Price-to-Book (P/B)

$$\text{PB} = \frac{\text{Stock Price}}{\text{Book Value per Share}} \quad (\text{B.15})$$

Earnings Yield

$$\text{Earnings Yield} = \frac{\text{EPS}}{\text{Stock Price}} \quad (\text{B.16})$$

EV/EBITDA

$$\text{EV/EBITDA} = \frac{\text{Enterprise Value}}{\text{EBITDA}} \quad (\text{B.17})$$

Growth Metrics

Revenue Growth

$$\text{Revenue Growth} = \frac{\text{Revenue}_t - \text{Revenue}_{t-1}}{\text{Revenue}_{t-1}} \quad (\text{B.18})$$

Earnings Growth

$$\text{Earnings Growth} = \frac{\text{EPS}_t - \text{EPS}_{t-1}}{\text{EPS}_{t-1}} \quad (\text{B.19})$$

B.2 | Technical Indicators

MACD (Moving Average Convergence Divergence)

$$\text{MACD} = \text{EMA}_{\text{short}} - \text{EMA}_{\text{long}} \quad (\text{B.20})$$

RSI (Relative Strength Index)

$$\text{RSI} = 100 - \frac{100}{1 + \text{RS}}, \quad \text{RS} = \frac{\text{Avg Gain}}{\text{Avg Loss}} \quad (\text{B.21})$$

VWAP (Volume Weighted Average Price)

$$\text{VWAP}_t = \frac{\sum_{i=1}^t (P_i \cdot V_i)}{\sum_{i=1}^t V_i} \quad (\text{B.22})$$

Z-Score

$$Z = \frac{P_t - \bar{P}}{\sigma} \quad (\text{B.23})$$

Beta (Systematic Risk)

$$\beta = \frac{\text{Cov}(R_s, R_m)}{\text{Var}(R_m)} \quad (\text{B.24})$$

B.3 | Volatility and Derivatives-Based Metrics

Put Skew Ratio

$$\text{ATM Skew} = \frac{\text{IV}_{\text{Put, ATM}} - \text{IV}_{\text{Call, ATM}}}{\text{IV}_{\text{Call, ATM}}} \quad (\text{B.25})$$

where:

- $\text{IV}_{\text{Put, ATM}}$ is the implied volatility of at-the-money put options,
- $\text{IV}_{\text{Call, ATM}}$ is the implied volatility of at-the-money call options.

$$\text{Put Skew Ratio} = \frac{IV_{\text{Put, ATM}}}{IV_{\text{Call, ATM}}} \quad (\text{B.26})$$

Implied Volatility – Black-Scholes Model

$$C = SN(d_1) - Ke^{-rt}N(d_2) \quad (\text{B.27})$$

$$d_1 = \frac{\ln\left(\frac{S}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)t}{\sigma\sqrt{t}}, \quad d_2 = d_1 - \sigma\sqrt{t} \quad (\text{B.28})$$

Historical Volatility

$$\text{HV} = \sqrt{\frac{1}{n-1} \sum_{t=1}^n (r_t - \bar{r})^2}, \quad r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \quad (\text{B.29})$$

ATR (Average True Range)

$$\text{ATR}_t = \text{EMA}_n(\text{True Range}_t) \quad (\text{B.30})$$

where True Range is defined as:

$$\max\{H_t - L_t, |H_t - C_{t-1}|, |L_t - C_{t-1}|\}$$

Prompts

This appendix contains all prompts used or tuned during Experiment 1.

C.1 | Baseline Strategy Prompt

```
1 User_Context:
2   Stock_Data:
3     Price:
4       Open: "{Open}"
5       High: "{High}"
6       Low: "{Low}"
7       Close: "{Close}"
8       Volume: "{Volume}"
9
10    Technical_Analysis:
11      Moving_Averages:
12        MA_Values:
13          20MA: "{20MA}"
14          50MA: "{50MA}"
15          100MA: "{100MA}"
16          200MA: "{200MA}"
17
18        MACD:
19          Value: "{MACD}"
20          Signal_Line: "{Signal_Line}"
21
22        RSI:
23          Value: "{RSI}"
24
25        ATR: "{ATR}"
26
27 System_Context:
28   persona: {persona}
29   risk_profile: {risk_profile}
30   portfolio_objectives: {portfolio_objectives}
31
32 Instructions:
33   - Evaluate the stock price data (Open, High, Low, Close, Volume) alongside
34     technical indicators (MACD, RSI, Bollinger Bands, ATR, and Moving Averages).
35   - Use both price and indicators collectively to make a trade decision
36     consistent with the risk profile, and portfolio objectives. Either LONG or
```

```
33     SHORT.  
34 Output:  
35     action: Str. LONG or SHORT.
```

C.2 | Writer-Trainer Prompt

```

1 User_Context:
2   Stock_Data:
3     General:
4       Quarter_Number: "{Quarter}"
5       Week_Number: "{week_number}"
6       Month_Number: "{month_number}"
7       Beta: {Market_Beta}
8       Classification: {classification}
9
10    Price:
11      Open: "{Open}"
12      High: "{High}"
13      Low: "{Low}"
14      Close: "{Close}"
15      Volume: "{Volume}"
16      Weekly_Past_Returns: {Weekly_Past_Returns}
17      Stock_Cumulative_Return: {Daily_Cumulative_Return}
18
19    Historical_Volatility:
20      HV_Open: "{HV_Open}"
21      HV_High: "{HV_High}"
22      HV_Low: "{HV_Low}"
23      HV_Close: "{HV_Close}"
24      HV_Volume: "{HV_Volume}"
25
26    Implied_Volatility:
27      IV_Open: "{IV_Open}"
28      IV_High: "{IV_High}"
29      IV_Low: "{IV_Low}"
30      IV_Close: "{IV_Close}"
31
32    Fundamental_Data:
33      Financials:
34        Net_Income: "{Net_Income}"
35        Invested_Capital: "{Invested_Capital}"
36        Total_Assets: "{Total_Assets}"
37        Return_on_Investment: "{Return_on_Investment}"
38        Shareholder_Equity: "{Shareholder's_Equity}"
39      Ratios:
40        Current_Ratio: "{Current_Ratio}"
41        Quick_Ratio: "{Quick_Ratio}"
42        Debt_to_Equity_Ratio: "{Debt_to_Equity_Ratio}"
43        Long_Term_Debt_to_Capital: "{Long_term_Debt_/Capital}"
44        Price_to_Book_Ratio: "{Price_to_Book_Ratio}"
45        Price_to_Sales_Ratio: "{Price_to_Sales_Ratio}"
46        Price_to_FCF_Ratio: "{Price_to_FCF_Ratio}"
47        PE_Ratio: "{PE_Ratio}"
48        Return_on_Assets: "{Return_on_Assets}"
49        Return_on_Equity: "{Return_on_Equity}"
50    Margins:

```

```

51     Gross_Margin: "{Gross_Margin}"
52     Operating_Margin: "{Operating_Margin}"
53     EBIT_Margin: "{EBIT_Margin}"
54     EBITDA_Margin: "{EBITDA_Margin}"
55     Pre_Tax_Profit_Margin: "{Pre_Tax_Profit_Margin}"
56     Net_Profit_Margin: "{Net_Profit_Margin}"
57 Other_Metrics:
58     EPS: "{EPS}"
59     TTM_Net_EPS: "{TTM_Net_EPS}"
60     Free_Cash_Flow_Per_Share: "{Free_Cash_Flow_Per_Share}"
61     Operating_Cash_Flow_Per_Share: "{Operating_Cash_Flow_Per_Share}"
62     Book_Value_Per_Share: "{Book_Value_per_Share}"
63     Asset_Turnover: "{Asset_Turnover}"
64     Inventory_Turnover_Ratio: "{Inventory_Turnover_Ratio}"
65     Days_Sales_In_Receivables: "{Days_Sales_In_Receivables}"
66 QoQ_Growth:
67     Quick_Ratio: "{Quick_Ratio_QoQ_Growth}"
68     Current_Ratio: "{Current_Ratio_QoQ_Growth}"
69     Debt_to_Equity_Ratio: "{Debt_to_Equity_Ratio_QoQ_Growth}"
70     Gross_Margin: "{Gross_Margin_QoQ_Growth}"
71     Operating_Margin: "{Operating_Margin_QoQ_Growth}"
72     EBIT_Margin: "{EBIT_Margin_QoQ_Growth}"
73     Net_Profit_Margin: "{Net_Profit_Margin_QoQ_Growth}"
74     Asset_Turnover: "{Asset_Turnover_QoQ_Growth}"
75     Inventory_Turnover_Ratio: "{Inventory_Turnover_Ratio_QoQ_Growth}"
76     Price_to_Book_Ratio: "{Price_to_Book_Ratio_QoQ_Growth}"
77     PE_Ratio: "{PE_Ratio_QoQ_Growth}"
78     EPS: "{EPS_QoQ_Growth}"
79     Net_Income: "{Net_Income_QoQ_Growth}"
80     Free_Cash_Flow_Per_Share: "{Free_Cash_Flow_Per_Share_QoQ_Growth}"
81     Operating_Cash_Flow_Per_Share: "{Operating_Cash_Flow_Per_Share_QoQ_Growth}"
82     "
83     Return_on_Equity: "{Return_on_Equity_QoQ_Growth}"
84     Return_on_Assets: "{Return_on_Assets_QoQ_Growth}"
85 YoY_Growth:
86     Quick_Ratio: "{Quick_Ratio_YoY_Growth}"
87     Current_Ratio: "{Current_Ratio_YoY_Growth}"
88     Debt_to_Equity_Ratio: "{Debt_to_Equity_Ratio_YoY_Growth}"
89     Gross_Margin: "{Gross_Margin_YoY_Growth}"
90     Operating_Margin: "{Operating_Margin_YoY_Growth}"
91     EBIT_Margin: "{EBIT_Margin_YoY_Growth}"
92     Net_Profit_Margin: "{Net_Profit_Margin_YoY_Growth}"
93     Asset_Turnover: "{Asset_Turnover_YoY_Growth}"
94     Inventory_Turnover_Ratio: "{Inventory_Turnover_Ratio_YoY_Growth}"
95     Price_to_Book_Ratio: "{Price_to_Book_Ratio_YoY_Growth}"
96     PE_Ratio: "{PE_Ratio_YoY_Growth}"
97     EPS: "{EPS_YoY_Growth}"
98     Net_Income: "{Net_Income_YoY_Growth}"
99     Free_Cash_Flow_Per_Share: "{Free_Cash_Flow_Per_Share_YoY_Growth}"
100    Operating_Cash_Flow_Per_Share: "{Operating_Cash_Flow_Per_Share_YoY_Growth}"
101    "
102    Return_on_Equity: "{Return_on_Equity_YoY_Growth}"

```

```

101     Return_on_Assets: "{Return_on_Assets_YoY_Growth}"
102
103     Technical_Analysis:
104         Moving_Averages:
105             MA_Values:
106                 20MA: "{20MA}"
107                 50MA: "{50MA}"
108                 100MA: "{100MA}"
109                 200MA: "{200MA}"
110             MA_Slopes:
111                 20MA_Slope: "{20MA_Slope}"
112                 50MA_Slope: "{50MA_Slope}"
113                 100MA_Slope: "{100MA_Slope}"
114                 200MA_Slope: "{200MA_Slope}"
115             MA_Z_Scores:
116                 20MA_Z: "{20MA_Z}"
117                 50MA_Z: "{50MA_Z}"
118                 100MA_Z: "{100MA_Z}"
119                 200MA_Z: "{200MA_Z}"
120         MACD:
121             Value: "{MACD}"
122             Signal_Line: "{Signal_Line}"
123             MACD_Strength: {MACD_Strength}
124             MACD_Strength_Slope: {MACD_Strength_Slope}
125         RSI:
126             Value: "{RSI}"
127             RSI_Divergence: {RSI_Divergence}
128         Bollinger:
129             Bollinger_Bands_Width: {BB_Width}
130         ATR: "{ATR}"
131
132     Macro_Data:
133         Macro_Indices:
134             SPX:
135                 Close: "{SPX_Close}"
136                 Close_MA: "{SPX_Close_MA}"
137                 Close_Slope: "{SPX_Close_Slope}"
138                 Close_Z: "{SPX_Close_Z}"
139             NDX:
140                 Close: "{NDX_Close}"
141                 Close_MA: "{NDX_Close_MA}"
142                 Close_Slope: "{NDX_Close_Slope}"
143                 Close_Z: "{NDX_Close_Z}"
144             VIX:
145                 Close: "{VIX_Close}"
146                 Close_MA: "{VIX_Close_MA}"
147                 Close_Slope: "{VIX_Close_Slope}"
148                 Close_Z: "{VIX_Close_Z}"
149         Bonds:
150             TNX_Close: "{TNX_Close}"
151             TNX_Close_MA: "{TNX_Close_MA}"
152             TNX_Close_Slope: "{TNX_Close_Slope}"

```

```

153     TNX_Close_Z: "{TNX_Close_Z}"
154
155     Economic_Data:
156         PMI: "{PMI}"
157         Retail_Sales: "{Retail_Sales}"
158         Employment: "{Employment}"
159         Treasury_Yields: "{Treasury_Yields}"
160         Consumer_Confidence: "{Consumer_Confidence}"
161         Housing_Starts: "{Housing_Starts}"
162         GDP: "{GDP}"
163         Yield_Curve: "{Yield_Curve}"
164         YoY_Growth:
165             PMI_YoY: "{PMI_YoY}"
166             Retail_Sales_YoY: "{Retail_Sales_YoY}"
167             Employment_YoY: "{Employment_YoY}"
168             Treasury_Yields_YoY: "{Treasury_Yields_YoY}"
169             Consumer_Confidence_YoY: "{Consumer_Confidence_YoY}"
170             Housing_Starts_YoY: "{Housing_Starts_YoY}"
171             GDP_YoY: "{GDP_YoY}"
172             Yield_Curve_YoY: "{Yield_Curve_YoY}"
173         QoQ_Growth:
174             PMI_QoQ: "{PMI_QoQ}"
175             Retail_Sales_QoQ: "{Retail_Sales_QoQ}"
176             Employment_QoQ: "{Employment_QoQ}"
177             Treasury_Yields_QoQ: "{Treasury_Yields_QoQ}"
178             Consumer_Confidence_QoQ: "{Consumer_Confidence_QoQ}"
179             Housing_Starts_QoQ: "{Housing_Starts_QoQ}"
180             GDP_QoQ: "{GDP_QoQ}"
181             Yield_Curve_QoQ: "{Yield_Curve_QoQ}"
182
183     Options_Data:
184         OTM_IV:
185             Call: "{OTM_IV_Call}"
186             Put: "{OTM_IV_Put}"
187             Put_IV_Skews: "{OTM_Skew}"
188         MA:
189             Call: "{MA_OTM_IV_Call}"
190             Put: "{MA_OTM_IV_Put}"
191             Put_IV_Skews: "{MA_OTM_Skew}"
192         ATM_IV:
193             Call: "{ATM_IV_Call}"
194             Put: "{ATM_IV_Put}"
195             Put_IV_Skews: "{ATM_Skew}"
196         MA:
197             Call: "{MA_ATM_IV_Call}"
198             Put: "{MA_ATM_IV_Put}"
199             Put_IV_Skews: "{MA_ATM_Skew}"
200         ITM_IV:
201             Call: "{ITM_IV_Call}"
202             Put: "{ITM_IV_Put}"
203             Put_IV_Skews: "{ITM_Skew}"
204         MA:

```

```

205     Call: "{MA_ITM_IV_Call}"
206     Put: "{MA_ITM_IV_Put}"
207     Put_IV_Skews: "{MA_ITM_Skew}"
208
209 Expert_Decision:
210     Expert_Action: {Expert_Action}
211     Next_Day_Returns: {Next_Day_Returns}
212     Next_Week_Returns: {Next_Week_Returns}
213     Next_Month_Returns: {Next_Month_Returns}
214
215 System_Context:
216     persona: You are an ICL Exemplar Generator, creating samples to train a prompt
217     generator.
218     risk_profile: {risk_profile}
219     portfolio_objectives: {portfolio_objectives}
220     instructions: |
221         Deduce the rationale behind the expert decision by thoroughly analyzing the
222         provided dense `context`. Focus on identifying impactful patterns, reasoning
223         pathways, and interactions across categories while aligning decisions with
224         `portfolio_objectives` and the specified `risk_profile`. Ensure your
225         analysis builds insights for future iterations.
226
227 Process:
228     1. Comprehensive Data Analysis:
229         - Stock Data: Analyze recent trends in price, volume, and volatility
230         metrics (HV, IV) over the last month. Incorporate technical indicators such
231         as RSI and moving averages (e.g., 20MA, 50MA, 200MA) to evaluate momentum
232         and trend strength.
233         - Macro Data: Evaluate indices (e.g., SPX, VIX), treasury yields, and
234         yield curves to assess market sentiment and risk-on/risk-off dynamics.
235         - Economic Data: Examine leading indicators (e.g., GDP, PMI, Retail
236         Sales) for potential macroeconomic impacts on market performance. Prioritize
237         quarterly or monthly trends over annual data.
238         - Fundamental Data: Assess ratios (e.g., PE Ratio, Debt-to-Equity),
239         growth metrics (e.g., EPS YoY), and margins (e.g., Operating Margin) to
240         evaluate company performance and risk.
241         - Options Data: Analyze implied volatility levels, skews, and option
242         demand shifts (e.g., IV_Skews) to infer market sentiment.
243         - Temporal and Seasonal Factors: are the moves seasonal, depending on
244         the `month_number` or `quarter_number`.
245         - Highlight cross-category relationships: Draw causal links between
246         categories, such as how macro shifts (e.g., rising VIX) influence stock
247         momentum or how IV skews correlate with beta-driven performance.
248         - Avoid referencing fields prefixed with "Expert_" or "Next_" under `
249         Expert_Decision`.
250
251     2. Logical and Structured Reasoning (CoT):
252         - Construct a clear and logical explanation connecting data across
253         categories. Emphasize cause-and-effect relationships and highlight
254         interdependencies.
255         - Provide examples by referencing specific metrics and their causal
256         impacts, such as "A rising RSI combined with falling IV skew indicates

```

stronger bullish sentiment."

236

237 3. Feature-Based Explanation and Ranking:

238 - Identify and rank 6 to 15 impactful features by their influence on
the strategy, spanning multiple data categories.

239 - Ensure at least one feature is selected from each of the following
groups: 'Stock_Data', 'Options_Data', 'Macro_Data', 'Economic_Data', '
Fundamental_Data', 'Technical_Analysis'.

240 - Prioritize metrics directly supporting the decision and demonstrate
their interactions. For example, explain how "RSI divergence with falling
VIX supports bullish sentiment."

241 - Use a Likert scale (1 to 3) to rank each feature by importance:

242 - 1: Minimal relevance; can be excluded.

243 - 2: Moderately relevant; contributes but isn't crucial.

244 - 3: Highly relevant; key driver of the decision.

245

246 4. Reflective Insight Generation:

247 - Critique the current decision-making process by identifying strengths
, weaknesses, and potential biases.

248 - Reflect on how external factors (e.g., macroeconomic shocks, policy
changes) or misjudged metrics impacted the strategy.

249 - Incorporate learnings to improve subsequent prompts, focusing on
feature prioritization and cross-category insights.

250

251 Example Explanations:

252 - "The RSI (importance of 2) crossing above 70 indicates overbought
conditions, historically signaling a reversal, especially when combined with
high implied volatility."

253 - "A widening yield curve paired with a 2% YoY increase in GDP signals
improved economic growth, favoring cyclicals like industrials."

254 - "A drop in the debt-to-equity ratio (importance of 3) by 10% reduces
leverage risks, aligning with the portfolio's low-risk objectives."

255

256 Deliver:

257 - explanation: Concise analyses (max 350 words), with your structured
reasoning and aligning to feature ranking in 'features', detailing a review
of the expert's rationale, focusing on alignment with objectives and
reasoning from your reflection.

258 - features: Top 6 feature names that directly support the decision, and
ranked from highest importance to lowest. List features as they are named in
the context using dot notation, don't paraphrase, e.g. 'Options_Data.OTM_IV
.Call', 'Economic_Data.PMI'

C.3 | Writer-Generator Prompt

```

1 User_Context:
2   Features_Reranked:
3     {{ features_reranked_dict }}
4
5   Judge_Feedback: |
6     {{ judge_critique }}
7
8   risk_profile: "{{ risk_profile }}"
9   portfolio_objectives: "{{ portfolio_objectives }}"
10
11 System_Context:
12   persona: "You are an LLM prompt engineer generating updated instructions for a
13     trading strategy generator. Your job is to translate expert critique and
14     reweighted features into structured, causal, and risk-aware prompting
15     guidance."
16
17   instructions: |
18     Use the `Judge_Feedback` and `Features_Reranked` to write a revised `
19     instructions:` block for a strategy prompt.
20     The result will be consumed by another LLM tasked with generating a
21     directional trading recommendation and rationale.
22
23     Your output must:
24     - Reflect the feature priorities provided in `Features_Reranked`.
25     - Incorporate feedback and reasoning adjustments from `Judge_Feedback`.
26     - Integrate feature causality and cross-category relationships.
27     - Be customized to the provided `risk_profile` and `portfolio_objectives`.
28     - Guide the next model to produce an `action` (LONG or SHORT), a causal `
29     explanation`, and a weighted feature list (Likert 1 to 3 scale).
30
31     Make your output structured, focused, and compliant with the expected prompt
32     formatting.
33
34 Deliver:
35   instructions: |
36     Develop a trading strategy for the next month using the provided financial,
37     technical, macroeconomic, and options data. Your recommendation must align
38     with the `portfolio_objectives` and `risk_profile` stated.
39
40   Based on Judge feedback:
41     {{ judge_critique }}
42
43   Prioritize the following features (ranked by importance):
44     {% for feature, weight in Features_Reranked.items() %}
45     - {{ feature }}, Weight {{ weight }}
46     {% endfor %}
47
48   Process:
49     1. Evaluate data across stock price, technical indicators, fundamental
50     ratios, options sentiment, and macroeconomic trends.

```

```
40 2. Reflect on prior strategy failures and improve causal assumptions
    accordingly.
41 3. Use 5 to 7 of the ranked features in your final decision.
42 4. Prefer multi-category causal interactions: e.g., RSI + IV skew + VIX
    slope.
43
44 Strategy rules by risk profile:
45 {% if risk_profile == 'high' %}
46 - Emphasize momentum and volatility signals (e.g., MACD, ATR, IV).
47 - Support risk-on moves with confirming macro or options signals.
48 {% elif risk_profile == 'low' %}
49 - Emphasize stability and valuation (e.g., Operating Margin, Debt Ratios,
    ITM skew).
50 - Use macroeconomic stability (e.g., Consumer Confidence, low VIX) as
    confirmation.
51 {% endif %}
52
53 Output must include:
54 - `action`: LONG or SHORT
55 - `explanation`: Causal logic <= 350 words using ranked features
56 - Weighted features list in the format:
57   - Options_Data.ATM_IV.Call, Weight 3
58   - Macro_Data.VIX.Close_Slope, Weight 3
59   - Technical_Analysis.RSI.Value, Weight 2
```

C.4 | Judge Prompt v1

```

1 User_Context (User):
2   Strategy_Template: |
3     "{{ train_template }}"
4   Tuning_History: |
5     "{{ shared_memory }}"
6
7 System_Context:
8   persona: "You are a prompt judge evaluating trade strategy templates."
9
10  instructions: |
11    Evaluate the given 'Strategy_Template' and its 'Tuning_History'. Your goal
12    is to critique the quality, coherence, and causal alignment of the strategy
13    and return two things:
14
15    1. A ranked list of features (from the current template) that should guide
16    the next version, based on current market regime and strategic coherence.
17
18    2. A concise but comprehensive paragraph explaining what went wrong or right
19    , what assumptions must be corrected, and how feature relevance should shift
20    . This will condition the next prompt instructions.
21
22  Steps:
23
24    - Identify mismatches between assumptions and results (e.g., over-reliance
25    on RSI in volatility spikes, or ignored macro drawdowns).
26
27    - Prioritize features that had consistent explanatory value and discard
28    redundant or low-informative signals.
29
30    - Reflect on short-, mid-, and long-term causal patterns across Stock, Macro
31    , Technical, Options, and Fundamentals.
32
33    - Emphasize regime alignment: bullish, bearish, or volatile.
34
35    - Provide Likert-weighted features, 1 to 3.
36
37 Deliver:
38   features: >
39     A comma-separated list of features, ordered from highest to lowest
40     importance, drawn only from the existing Strategy_Template.
41     Format: "Macro_Data.VIX.Close_Slope, Options_Data.ATM_IV.Call,
42     Technical_Analysis.RSI.Value"
43
44   judge_critique: >
45     A 1 to 2 paragraph explanation that summarizes:
46     - What worked or failed in the current strategy.
47     - Which features were under- or over-weighted.
48     - What macro or volatility conditions altered the validity of certain
49     assumptions.
50     - How feature interactions (e.g., IV skew + GDP growth) should inform the
51     next prompt.
52     - Use clean causal language and tie feature weights to their relevance in
53     the next strategy.

```

C.5 | Strategy Prompt v1: Tuned Prompt

```
1 User_Context:
2   Stock_Data:
3     General:
4       Beta: {Market_Beta}
5       Classification: {classification}
6
7     Last_Weeks_Price:
8       Close: "{Close}"
9       Volume: "{Volume}"
10
11     Weekly_Past>Returns: "{Weekly_Past>Returns}"
12
13     Historical_Volatility:
14       HV_Close: "{HV_Close}"
15
16     Implied_Volatility:
17       IV_Close: "{IV_Close}"
18
19   Fundamental_Data:
20     Ratios:
21       Current_Ratio: "{Current_Ratio}"
22       Quick_Ratio: "{Quick_Ratio}"
23       Debt_to_Equity_Ratio: "{Debt_to_Equity_Ratio}"
24       PE_Ratio: "{PE_Ratio}"
25     Margins:
26       Gross_Margin: "{Gross_Margin}"
27       Operating_Margin: "{Operating_Margin}"
28       Net_Profit_Margin: "{Net_Profit_Margin}"
29     Growth Metrics:
30       EPS_YoY: "{EPS_YoY_Growth}"
31       Net_Income_YoY: "{Net_Income_YoY_Growth}"
32       Free_Cash_Flow_YoY: "{Free_Cash_Flow_Per_Share_YoY_Growth}"
33
34   Technical_Analysis:
35     Moving_Averages:
36       20MA: "{20MA}"
37       50MA: "{50MA}"
38       200MA: "{200MA}"
39     MACD:
40       Value: "{MACD}"
41       Signal_Line: "{Signal_Line}"
42       MACD_Strength: {MACD_Strength}
43     RSI:
44       Value: "{RSI}"
45     ATR: "{ATR}"
46
47   Macro_Data:
48     Macro_Indices:
49       SPX:
50         Close: "{SPX_Close}"
```

```

51     Close_20MA: "{SPX_Close_MA}"
52     Close_Slope: "{SPX_Close_Slope}"
53     VIX:
54         Close: "{VIX_Close}"
55         Close_20MA: "{VIX_Close_MA}"
56         Close_Slope: "{VIX_Close_Slope}"
57     Economic_Data:
58         GDP_QoQ: "{GDP_QoQ}"
59         PMI: "{PMI}"
60         Consumer_Confidence_QoQ: "{Consumer_Confidence_QoQ}"
61         M2_Money_Supply_QoQ: "{M2_Money_Supply_QoQ}"
62         PPI_YoY: "{PPI_YoY}"
63         Treasury_Yields_YoY: "{Treasury_Yields_YoY}"
64
65     Options_Data:
66         Put_IV_Skews:
67             OTM_Skew: "{OTM_Skew}"
68             ATM_Skew: "{ATM_Skew}"
69             ITM_Skew: "{ITM_Skew}"
70         20Day_Moving_Averages:
71             OTM_Skew_MA: "{MA_OTM_Skew}"
72             ATM_Skew_MA: "{MA_ATM_Skew}"
73             ITM_Skew_MA: "{MA_ITM_Skew}"
74
75     System_Context:
76         persona: {persona}
77         risk_profile: {risk_profile}
78         portfolio_objectives: {portfolio_objectives}
79         instructions: |
80             Develop a trading strategy for the next month based on the given context and
            aligned with the specified 'portfolio_objectives' and 'risk_profile'. Use
            the following process:
81
82             1. Comprehensive Data Analysis:
83                 - Stock Data: Examine price trends, volume, and HV/IV metrics for
            momentum or risk signals.
84                 - Fundamental Data: Focus on profitability margins and ratios based on
            risk tolerance.
85                 - Technical Analysis: Use RSI for overbought/oversold conditions, MAs for
            trend confirmation, and MACD for momentum analysis.
86                 - Macro Data: Evaluate GDP, PMI, and VIX trends to assess broader
            sentiment.
87                 - Options Data: Prioritize implied volatility metrics to capture
            sentiment shifts.
88                 - Dynamic Feature Weighting by Risk Profile:
89                     - High-Risk Profile:
90                         - Prioritize volatility and momentum indicators such as RSI, MACD, ATR
            , and Options_Data.ATM_IV.Call.
91                         - Weigh macroeconomic indicators (e.g., VIX, GDP_QoQ) for risk-on
            sentiment.
92                     - Low-Risk Profile:

```

```
93         - Focus on stability metrics like Debt-to-Equity Ratio, Operating
          Margin, and Current Ratio.
94         - Analyze implied volatility skews (Options_Data.ITM_Skew) for
          downside risk mitigation.
95         - Use macroeconomic stability indicators such as Consumer Confidence
          QoQ and Treasury Yields.
96
97     2. Scenario-Based Validation:
98         - Simulate contrasting market conditions (bullish, bearish, ranging).
99         - Validate feature sensitivity under external shocks (e.g., rising VIX or
          GDP declines).
100        - Propose feature adjustments to improve adaptability and robustness.
101
102     3. Feature Ranking and Aggregation:
103        - Rank features dynamically using a Likert scale (1: Minimal, 3: Critical
          ).
104        - Aggregate consistent high-ranking features across scenarios to ensure
          robust decision-making.
105
106     Output:
107         action: str. LONG or SHORT.
108         explanation: String. A clear, concise rationale (max 350 words) including
          the top 5 weighted features with the news as a factor used in decision-
          making (ICL Example; "Stock_Data.Price.Close, Weight 3, Technical_Analysis.
          RSI.Value, Weight 1, Options_Data.ATM_IV.Call, Weight 2"), and if `
          news_factors` was provided, the top 3 ranked news factors, weighted the same
          using the Likert scale (ICL Example; "Earning Call next month, with
          positive analyst expectations, Weight 3").
```

C.6 | Strategy Prompt v2: In-Context Memory Prompt

```

1 User_Context:
2   Last_Strategy_Used_Data:
3     Cumulative>Returns: "{Last_LLM_Strat_Cum>Returns}"
4     Peak>Returns: "{Last_LLM_Strat_Best>Returns}"
5     Worst_Losses: "{Last_LLM_Strat_Worse>Returns}"
6     Rationale: |
7       "{Last_LLM_Strat}"
8
9   Stock_Data:
10    General:
11      Beta: {Market_Beta}
12      Classification: {classification}
13
14    Last_Weeks_Price:
15      Close: "{Close}"
16      Volume: "{Volume}"
17
18    Weekly_Past>Returns: "{Weekly_Past>Returns}"
19
20    Historical_Volatility:
21      HV_Close: "{HV_Close}"
22
23    Implied_Volatility:
24      IV_Close: "{IV_Close}"
25
26    Fundamental_Data:
27      Ratios:
28        Current_Ratio: "{Current_Ratio}"
29        Quick_Ratio: "{Quick_Ratio}"
30        Debt_to_Equity_Ratio: "{Debt_to_Equity_Ratio}"
31        PE_Ratio: "{PE_Ratio}"
32      Margins:
33        Gross_Margin: "{Gross_Margin}"
34        Operating_Margin: "{Operating_Margin}"
35        Net_Profit_Margin: "{Net_Profit_Margin}"
36      Growth Metrics:
37        EPS_YoY: "{EPS_YoY_Growth}"
38        Net_Income_YoY: "{Net_Income_YoY_Growth}"
39        Free_Cash_Flow_YoY: "{Free_Cash_Flow_Per_Share_YoY_Growth}"
40
41    Technical_Analysis:
42      Moving_Averages:
43        20MA: "{20MA}"
44        50MA: "{50MA}"
45        200MA: "{200MA}"
46      MACD:
47        Value: "{MACD}"
48        Signal_Line: "{Signal_Line}"
49        MACD_Strength: {MACD_Strength}
50      RSI:

```

```

51     Value: "{RSI}"
52     ATR: "{ATR}"
53
54     Macro_Data:
55         Macro_Indices:
56             SPX:
57                 Close: "{SPX_Close}"
58                 Close_20MA: "{SPX_Close_MA}"
59                 Close_Slope: "{SPX_Close_Slope}"
60             VIX:
61                 Close: "{VIX_Close}"
62                 Close_20MA: "{VIX_Close_MA}"
63                 Close_Slope: "{VIX_Close_Slope}"
64         Economic_Data:
65             GDP_QoQ: "{GDP_QoQ}"
66             PMI: "{PMI}"
67             Consumer_Confidence_QoQ: "{Consumer_Confidence_QoQ}"
68             M2_Money_Supply_QoQ: "{M2_Money_Supply_QoQ}"
69             PPI_YoY: "{PPI_YoY}"
70             Treasury_Yields_YoY: "{Treasury_Yields_YoY}"
71
72     Options_Data:
73         Put_IV_Skews:
74             OTM_Skew: "{OTM_Skew}"
75             ATM_Skew: "{ATM_Skew}"
76             ITM_Skew: "{ITM_Skew}"
77         20Day_Moving_Averages:
78             OTM_Skew_MA: "{MA_OTM_Skew}"
79             ATM_Skew_MA: "{MA_ATM_Skew}"
80             ITM_Skew_MA: "{MA_ITM_Skew}"
81
82     System_Context:
83     persona: {persona}
84     risk_profile: {risk_profile}
85     portfolio_objectives: {portfolio_objectives}
86     instructions: |
87         Develop a trading strategy for the next month that aligns with the
88         portfolio_objectives and risk_profile. Follow these steps:
89
90         1. Stock Analysis:
91             - Evaluate price trends: Compare the Close price against 20MA, 50MA, and
92             200MA to assess momentum or reversals.
93             - Analyze returns: Use Weekly Past Returns and Cumulative Returns to
94             validate trend sustainability.
95             - Contextualize volatility: Align HV_Close and HV_High with recent price
96             action for trend validation and risk management.
97             - Incorporate beta: Use beta to gauge sensitivity to market movements.
98             - High-Risk: Prioritize high-beta stocks with steep momentum surges.
99             - Low-Risk: Focus on low-volatility stocks with consistent volume.
100
101             - ICL Example: "Close price above 20MA and 50MA with steep 20MA slope
102             signals bullish momentum. Weekly returns confirm a sustainable uptrend."

```

98
99 2. Technical Analysis:
100 - Use RSI: Identify momentum signals (>70 overbought; <30 oversold) and
divergences for reversals.
101 - Validate with MACD: Use crossovers, slope, and MACD strength for
directional confidence.
102 - Monitor Bollinger Bands: Narrowing bands indicate volatility
compression; validate breakouts with volume spikes.
103 - High-Risk: Leverage RSI divergences, steep MA slopes, and Bollinger
Band contractions for breakouts.
104 - Low-Risk: Focus on stable MA patterns and Bollinger Band retractions.
105
106 - ICL Example: "RSI at 65, a positive MACD crossover, and Bollinger Band
contraction indicate bullish momentum."
107
108 3. Fundamental Analysis:
109 - Evaluate growth metrics: Use EPS YoY, Net Income YoY, and Free Cash
Flow YoY for profitability and sustainability.
110 - Prioritize ratios: Low Debt-to-Equity and high Current Ratios reflect
financial stability.
111 - High-Risk: Focus on aggressive growth metrics and earnings surprises.
112 - Low-Risk: Emphasize steady margins, cash flow, and moderate leverage.
113
114 - ICL Example: "EPS YoY growth of 25% and low Debt-to-Equity ratio of 0.5
support strong financial health, aligning with a LONG strategy."
115
116 4. Macro Analysis:
117 - Align with market sentiment:
118 - "SPX_Close_Slope > 0 && VIX_Close_Slope < 0" is Bullish (Risk-On)
119 - "SPX_Close_Slope < 0 && VIX_Close_Slope > 0" is Bearish (Risk-Off)
120 - Validate with economic indicators:
121 - "GDP_QoQ > 0 && PMI > 50" leads to Economic Expansion
122 - "M2_Growth > 8% && ISM_Prices_Paid > 50 && Durable_Goods > 0"
Increases Inflationary Pressures
123 - "Yield_Curve(10Y-2Y) < 0" Signals Recession Risk, especially if
already mentioned in 'Rationale'.
124
125 - ICL Examples:
126 - "SPX_Close_Slope > 0 && VIX_Close_Slope < 0 We have Market Confidence,
Favor Risk-On actions"
127 - "M2 Growth > 8% && Rising ISM Prices Paid we have an Inflationary
Environment"
128 - "GDP_QoQ Falling && PMI < 50 We have an Economic Slowdown, Reduce Risk
Exposure"
129
130 5. Options Analysis:
131 - Compare OTM, ATM, and ITM IV Skews: Assess differences to gauge market
sentiment and directional bias.
132 - High-Risk Strategies: Leverage IV spikes to capitalize on speculative
directional trades.
133 - Low-Risk Strategies: Focus on balanced IV levels and consistent HV for
steady, lower-risk sentiment plays.

```
134     - Example - ICL: "Rising ATM IV skew > 0, market pricing up move, with
135     stable HV supports a LONG position, as it indicates growing upside
136     expectations without excessive fear."
137
138     6. Reflection and Iterative Learning:
139     - If not None, evaluate 'Last_Strategy_Used_Data' for performance gaps
140     and reflect on your chosen 'action'.
141     - If not None, use CoT reasoning to adjust mismatches between past
142     assumptions and actual market behavior.
143     - Weigh the feature importance in your rationale, use a Likert scale of 3:
144     - level: 1
145       description: The feature has minimal relevance or impact; it is not
146       necessary and can be ignored.
147     - level: 2
148       description: The feature has some relevance and contributes to the
149       strategy but is not crucial.
150     - level: 3
151       description: The feature is important and significantly contributes
152       to achieving a successful strategy.
153
154     Output:
155     action: Str. LONG or SHORT.
156     explanation: String. A clear, concise rationale (max 350 words) including 3
157     weighted features used in decision-making (ICL Example; "Stock_Data.Price.
158     Close, Weight 3, Technical_Analysis.RSI.Value, Weight 1, Options_Data.ATM_IV
159     .Call, Weight 2").
```

C.7 | Strategy Prompt v3: Decomposed Prompting

```

1 User_Context:
2   Last_Strategy_Used_Data:
3     last_returns: "{Last_LLM_Strat>Returns}"
4     last_action: "{Last_LLM_Strat_Action}"
5     last_action_Confidence: "{Last_LLM_Strat_Action_Confidence}"
6     Rationale: |
7       ""{Last_LLM_Strat}""
8
9   Stock_Data:
10    General:
11      Beta: {Market_Beta}
12      Classification: {classification}
13
14    Last_Weeks_Price:
15      Close: "{Close}"
16      Volume: "{Volume}"
17
18    Weekly_Past>Returns: "{Weekly_Past>Returns}"
19
20    Historical_Volatility:
21      HV_Close: "{HV_Close}"
22
23    Implied_Volatility:
24      IV_Close: "{IV_Close}"
25
26    Fundamental_Data:
27      Ratios:
28        Current_Ratio: "{Current_Ratio}"
29        Quick_Ratio: "{Quick_Ratio}"
30        Debt_to_Equity_Ratio: "{Debt_to_Equity_Ratio}"
31        PE_Ratio: "{PE_Ratio}"
32      Margins:
33        Gross_Margin: "{Gross_Margin}"
34        Operating_Margin: "{Operating_Margin}"
35        Net_Profit_Margin: "{Net_Profit_Margin}"
36      Growth Metrics:
37        EPS_YoY: "{EPS_YoY_Growth}"
38        Net_Income_YoY: "{Net_Income_YoY_Growth}"
39        Free_Cash_Flow_YoY: "{Free_Cash_Flow_Per_Share_YoY_Growth}"
40
41    Technical_Analysis:
42      Moving_Averages:
43        20MA: "{20MA}"
44        50MA: "{50MA}"
45        200MA: "{200MA}"
46      MACD:
47        Value: "{MACD}"
48        Signal_Line: "{Signal_Line}"
49        MACD_Strength: {MACD_Strength}
50      RSI:

```

```

51     Value: "{RSI}"
52     ATR: "{ATR}"
53
54     Macro_Data:
55         Macro_Indices:
56             SPX:
57                 Close: "{SPX_Close}"
58                 Close_20MA: "{SPX_Close_MA}"
59                 Close_Slope: "{SPX_Close_Slope}"
60             VIX:
61                 Close: "{VIX_Close}"
62                 Close_20MA: "{VIX_Close_MA}"
63                 Close_Slope: "{VIX_Close_Slope}"
64         Economic_Data:
65             GDP_QoQ: "{GDP_QoQ}"
66             PMI: "{PMI}"
67             Consumer_Confidence_QoQ: "{Consumer_Confidence_QoQ}"
68             M2_Money_Supply_QoQ: "{M2_Money_Supply_QoQ}"
69             PPI_YoY: "{PPI_YoY}"
70             Treasury_Yields_YoY: "{Treasury_Yields_YoY}"
71
72     Options_Data:
73         Put_IV_Skews:
74             OTM_Skew: "{OTM_Skew}"
75             ATM_Skew: "{ATM_Skew}"
76             ITM_Skew: "{ITM_Skew}"
77         20Day_Moving_Averages:
78             OTM_Skew_MA: "{MA_OTM_Skew}"
79             ATM_Skew_MA: "{MA_ATM_Skew}"
80             ITM_Skew_MA: "{MA_ITM_Skew}"
81
82     System_Context(System):
83         Persona: {persona}
84         Risk_Profile: {risk_profile}
85         Portfolio_Objectives: {portfolio_objectives}
86         Instructions: |
87             Develop a LONG or SHORT trading strategy for a single stock only for the
            next Month that aligns with the 'portfolio_objectives' and 'risk_profile'.
            Follow these guidelines:
88
89             1. Stock Analysis:
90                 - Evaluate price trends: Compare the Close price against 20MA, 50MA, and
                200MA to assess momentum or reversals.
91                 - Analyze returns: Use Weekly Past Returns to validate trend
                sustainability.
92                 - Contextualize volatility: Align 'HV_Close' and 'HV_High' with recent
                price action for trend validation and risk management.
93                 - Incorporate beta: Use 'beta' to gauge sensitivity to market movements.
94                 - High-Risk: Leverage high-beta trend for momentum surges.
95                 - Low-Risk: Focus on low 'beta' and volatility.
96                 - ICL Example: "Close price above 20MA and 50MA with steep 20MA slope
                signals bullish momentum. Weekly returns confirm a sustainable uptrend."

```

```

97
98 2. Technical Analysis:
99   - Use RSI: Identify momentum signals (>70 overbought; <30 oversold) and
100     divergences for reversals.
101   - Validate with 'MACD': Use crossovers of 'MACD.Value' and 'Signal_Line',
102     and 'MACD_Strength' for directional confidence.
103   - High-Risk: Leverage 'RSI.value' divergences, and steep 'Moving_Averages
104     \ slopes.
105   - Low-Risk: Focus on stable 'Moving_Averages' patterns.
106   - ICL Example: "RSI at 65, a positive MACD crossover indicate bullish
107     momentum."
108
109 3. Fundamental Analysis:
110   - Evaluate growth metrics: Use 'EPS_YoY', 'Net_Income_YoY', and '
111     Free_Cash_Flow_YoY' for profitability and sustainability.
112   - Prioritize ratios: Low 'Debt_to_Equity_Ratio' and 'Current_Ratio'
113     reflect financial stability, especially in low risk confic
114   - High-Risk: Focus on aggressive 'Growth Metrics'.
115   - Low-Risk: Emphasize steady margins, cash flow, and moderate leverage.
116   - ICL Example: "EPS YoY growth of 25% and low Debt-to-Equity ratio of 0.5
117     support strong financial health, aligning with a LONG strategy."
118
119 4. Macro Analysis:
120   - Align with market sentiment across 'Macro_Data':
121     - "SPX_Close_Slope > 0 && VIX_Close_Slope < 0" is Bullish (Risk-On)
122     - "SPX_Close_Slope < 0 && VIX_Close_Slope > 0" is Bearish (Risk-Off)
123   - Validate with 'Economic_Data':
124     - "GDP_QoQ > 0 && PMI > 50" leads to Economic Expansion
125     - "M2_Growth > 8% && ISM_Prices_Paid > 50 && Durable_Goods > 0"
126     Increases Inflationary Pressures
127     - "Yield_Curve(10Y-2Y) < 0" Signals Recession Risk, especially if
128     already mentioned in 'Rationale'.
129
130   - ICL Examples:
131     - "SPX_Close_Slope > 0 && VIX_Close_Slope < 0 We have Market Confidence,
132       leverage high-risk actions for a risk profile"
133     - "M2 Growth > 8% && Rising ISM Prices Paid we have an Inflationary
134       Environment, low risk actions for low risk profile"
135     - "GDP_QoQ Falling && PMI < 50 We have an Economic Slowdown, low risk
136       trades for low risk profile, or high risk actions for high risk profiles"
137
138 5. Options Analysis:
139   - Compare 'OTM_Skew', 'ATM_Skew', and 'ITM_Skew' IV Skews: Assess
140     differences to gauge market sentiment and directional bias using their '20
141     Day_Moving_Averages'.
142   - High-Risk Strategies: Leverage IV spikes to capitalize on speculative
143     directional trades.
144   - Low-Risk Strategies: Focus on balanced IV levels and consistent HV for
145     steady, lower-risk trades.
146   - Example: "Rising ATM IV skew > 0, market pricing up move, with stable HV
147     supports a LONG position, as it indicates growing upside expectations
148     without excessive fear."

```

```

131
132 6. Performance Reflection and Strategic Adaptation:
133   - If 'Last_Strategy_Used_Data' is available:
134     - Outcome Evaluation:
135       - Assess the outcome of the previous strategy by examining '
136         last_returns' and the chosen 'last_action'.
137       - Determine if the result aligns with the expectations outlined in
138         the previous 'Rationale'.
139       - Identify if the direction (LONG or SHORT) led to desirable or
140         undesirable outcomes.
141       - Compare assumptions in the previous 'Rationale' with actual
142         outcomes and current 'User_Context'.
143       - Identify logical inconsistencies, misinterpreted signals, or
144         omitted influential features.
145       - Use chain-of-thought reasoning to explain the gap between
146         forecasted and observed behavior.
147     - Confidence Adjustment:
148       - Based on the observed gap and your reassessment, evaluate if '
149         last_action_Confidence' was accurate.
150       - Assign an updated Likert score (1 to 3) to your current '
151         action_confidence':
152         - 1: Low confidence; contradictory or weak alignment across
153           features.
154         - 2: Moderate confidence; partial alignment with moderate
155           evidence.
156         - 3: High confidence; strong convergence across key features.
157     - Risk-profile adjustment:
158       - If 'risk_profile' is "High-Risk", and your trade is high risk
159         high reward, consider increasing 'action_confidence' by +1, capped at 3.
160       - If 'risk_profile' is "Low-Risk" and downside risk is high,
161         consider decreasing 'action_confidence' by -1, floored at 1.
162     - Feature Attribution:
163       - Rank the importance of each major feature used in your current
164         rationale using a Likert scale (1 to 3):
165         - 1: Minimal contribution; not required for the decision.
166         - 2: Moderate contribution; relevant but not critical.
167         - 3: High contribution; pivotal to the trading decision.
168
169 Output:
170 action: Str. LONG or SHORT.
171 action_confidence: int. Likert scale (1-3) confidence in the proposed 'action
172   ', adjusted based on prior strategy outcome if 'Last_Strategy_Used_Data' is
173   available.
174 explanation: >
175   A concise rationale (max 350 words) justifying the proposed 'action'.
176   Include:
177     - The top 5 weighted features used in the decision, each labeled with its
178       Likert importance (1 to 3).
179     (e.g., "Stock_Data.Price.Close, Weight 3, Technical_Analysis.RSI.Value,
180       Weight 1, Options_Data.ATM_Skew, Weight 2")
181     - A reflective assessment of 'Last_Strategy_Used_Data', including:

```

- 165 - Whether the past `action` was successful and was it maintained given
prior `Rationale`.
- 166 - Whether prior confidence was justified and how this influenced the
current decision.

C.8 | Strategy Prompt v4: News Factors

```

1 User_Context:
2   Last_Strategy_Used_Data:
3     last_returns: "{Last_LLM_Strat>Returns}"
4     last_action: "{Last_LLM_Strat_Action}"
5     last_action_Confidence: "{Last_LLM_Strat_Action_Confidence}"
6     Rationale: |
7       ""{Last_LLM_Strat}""
8
9   Stock_Data:
10    General:
11      Beta: {Market_Beta}
12      Classification: {classification}
13
14    Last_Weeks_Price:
15      Close: "{Close}"
16      Volume: "{Volume}"
17
18    Weekly_Past>Returns: "{Weekly_Past>Returns}"
19
20    Historical_Volatility:
21      HV_Close: "{HV_Close}"
22
23    Implied_Volatility:
24      IV_Close: "{IV_Close}"
25
26    Fundamental_Data:
27      Ratios:
28        Current_Ratio: "{Current_Ratio}"
29        Quick_Ratio: "{Quick_Ratio}"
30        Debt_to_Equity_Ratio: "{Debt_to_Equity_Ratio}"
31        PE_Ratio: "{PE_Ratio}"
32      Margins:
33        Gross_Margin: "{Gross_Margin}"
34        Operating_Margin: "{Operating_Margin}"
35        Net_Profit_Margin: "{Net_Profit_Margin}"
36      Growth Metrics:
37        EPS_YoY: "{EPS_YoY_Growth}"
38        Net_Income_YoY: "{Net_Income_YoY_Growth}"
39        Free_Cash_Flow_YoY: "{Free_Cash_Flow_Per_Share_YoY_Growth}"
40
41    Technical_Analysis:
42      Moving_Averages:
43        20MA: "{20MA}"
44        50MA: "{50MA}"
45        200MA: "{200MA}"
46      MACD:
47        Value: "{MACD}"
48        Signal_Line: "{Signal_Line}"
49        MACD_Strength: {MACD_Strength}
50      RSI:

```

```

51     Value: "{RSI}"
52     ATR: "{ATR}"
53
54     Macro_Data:
55         Macro_Indices:
56             SPX:
57                 Close: "{SPX_Close}"
58                 Close_20MA: "{SPX_Close_MA}"
59                 Close_Slope: "{SPX_Close_Slope}"
60             VIX:
61                 Close: "{VIX_Close}"
62                 Close_20MA: "{VIX_Close_MA}"
63                 Close_Slope: "{VIX_Close_Slope}"
64         Economic_Data:
65             GDP_QoQ: "{GDP_QoQ}"
66             PMI: "{PMI}"
67             Consumer_Confidence_QoQ: "{Consumer_Confidence_QoQ}"
68             M2_Money_Supply_QoQ: "{M2_Money_Supply_QoQ}"
69             PPI_YoY: "{PPI_YoY}"
70             Treasury_Yields_YoY: "{Treasury_Yields_YoY}"
71
72     Options_Data:
73         Put_IV_Skews:
74             OTM_Skew: "{OTM_Skew}"
75             ATM_Skew: "{ATM_Skew}"
76             ITM_Skew: "{ITM_Skew}"
77         20Day_Moving_Averages:
78             OTM_Skew_MA: "{MA_OTM_Skew}"
79             ATM_Skew_MA: "{MA_ATM_Skew}"
80             ITM_Skew_MA: "{MA_ITM_Skew}"
81
82     news_factors: |
83         "{news_factors}"
84
85     System_Context (System):
86         Persona: {persona}
87         Risk_Profile: {risk_profile}
88         Portfolio_Objectives: {portfolio_objectives}
89         Instructions: |
90             Develop a LONG or SHORT trading strategy for a single stock only for the
             next Month that aligns with the 'portfolio_objectives' and 'risk_profile'.
             Follow these guidelines:
91
92             1. Stock Analysis:
93                 - Evaluate price trends: Compare the Close price against 20MA, 50MA, and
                 200MA to assess momentum or reversals.
94                 - Analyze returns: Use Weekly Past Returns to validate trend
                 sustainability.
95                 - Contextualize volatility: Align 'HV_Close' and 'HV_High' with recent
                 price action for trend validation and risk management.
96                 - Incorporate beta: Use 'beta' to gauge sensitivity to market movements.
97                 - High-Risk: Leverage high-beta trend for momentum surges.

```

```

98     - Low-Risk: Focus on low `beta` and volatility.
99     - ICL Example: "Close price above 20MA and 50MA with steep 20MA slope
100       signals bullish momentum. Weekly returns confirm a sustainable uptrend."
101
102 2. Technical Analysis:
103     - Use RSI: Identify momentum signals (>70 overbought; <30 oversold) and
104       divergences for reversals.
105     - Validate with `MACD`: Use crossovers of `MACD.Value` and `Signal_Line`,
106       and `MACD_Strength` for directional confidence.
107     - High-Risk: Leverage `RSI.value` divergences, and steep `Moving_Averages`
108       slopes.
109     - Low-Risk: Focus on stable `Moving_Averages` patterns.
110     - ICL Example: "RSI at 65, a positive MACD crossover indicate bullish
111       momentum."
112
113 3. Fundamental Analysis:
114     - Evaluate growth metrics: Use `EPS_YoY`, `Net_Income_YoY`, and `
115       Free_Cash_Flow_YoY` for profitability and sustainability.
116     - Prioritize ratios: Low `Debt_to_Equity_Ratio` and `Current_Ratio`
117       reflect financial stability, especially in low risk confic
118     - High-Risk: Focus on aggressive `Growth Metrics` and earnings news.
119     - Low-Risk: Emphasize steady margins, cash flow, and moderate leverage.
120     - ICL Example: "EPS YoY growth of 25% and low Debt-to-Equity ratio of 0.5
121       support strong financial health, aligning with a LONG strategy."
122
123 4. Macro Analysis:
124     - Align with market sentiment across `Macro_Data`:
125     - "SPX_Close_Slope > 0 && VIX_Close_Slope < 0" is Bullish (Risk-On)
126     - "SPX_Close_Slope < 0 && VIX_Close_Slope > 0" is Bearish (Risk-Off)
127     - Validate with `Economic_Data`:
128     - "GDP_QoQ > 0 && PMI > 50" leads to Economic Expansion
129     - "M2_Growth > 8% && ISM_Prices_Paid > 50 && Durable_Goods > 0"
130       Increases Inflationary Pressures
131     - "Yield_Curve(10Y-2Y) < 0" Signals Recession Risk, especially if
132       already mentioned in `Rationale`.
133
134     - ICL Examples:
135     - "SPX_Close_Slope > 0 && VIX_Close_Slope < 0 We have Market Confidence,
136       leverage high-risk actions for a risk profile"
137     - "M2 Growth > 8% && Rising ISM Prices Paid we have an Inflationary
138       Environment, low risk actions for low risk profile"
139     - "GDP_QoQ Falling && PMI < 50 We have an Economic Slowdown, low risk
140       trades for low risk profile, or high risk actions for high risk profiles"
141
142 5. Options Analysis:
143     - Compare `OTM_Skew`, `ATM_Skew`, and `ITM_Skew` IV Skews: Assess
144       differences to gauge market sentiment and directional bias using their `20
145       Day_Moving_Averages`.
146     - High-Risk Strategies: Leverage IV spikes to capitalize on speculative
147       directional trades.
148     - Low-Risk Strategies: Focus on balanced IV levels and consistent HV for
149       steady, lower-risk trades.

```

133 - Example: "Rising ATM IV skew > 0, market pricing up move, with stable HV
 134 supports a LONG position, as it indicates growing upside expectations
 135 without excessive fear."

134 6. News Analysis:

135 - Skip this if no news is provided in 'news_factors'.

136 - Factor in extracted news insights from 'news_factors' for directional
 137 confidence. Assign a "Market Impact Score" using a Likert Scale (1 to 3):

138 - 1: Minimal or no relevance.

139 - 3: Moderately relevant, contributes to the decision.

140 - 5: Strong influence or primary driver of trade rationale.

141 - Evaluate company announcements, corporate events, industry trends, and
 142 macro influences in tandem with technical and fundamental indicators.

142 - High-Risk: Prioritize sentiment-driven momentum plays.

143 - Low-Risk: Favor stability-backed moves corroborated by published
 144 market fundamentals.

144 - Example: "Positive earnings revision and analyst upgrades (news_factors)
 145 align with bullish technicals, reinforcing a LONG position. Impact score is
 146 3 out of 3."

145 7. Performance Reflection and Strategic Adaptation:

146 - If 'Last_Strategy_Used_Data' is available:

147 - Outcome Evaluation:

148 - Assess the outcome of the previous strategy by examining '
 149 last_returns' and the chosen 'last_action'.

150 - Determine if the result aligns with the expectations outlined in
 151 the previous 'Rationale'.

151 - Identify if the direction (LONG or SHORT) led to desirable or
 152 undesirable outcomes.

152 - Compare assumptions in the previous 'Rationale' with actual
 153 outcomes and current 'User_Context'.

153 - Identify logical inconsistencies, misinterpreted signals, or
 154 omitted influential features.

154 - Use chain-of-thought reasoning to explain the gap between
 155 forecasted and observed behavior.

155 - Confidence Adjustment:

156 - Based on the observed gap and your reassessment, evaluate if '
 157 last_action_Confidence' was accurate.

157 - Assign an updated Likert score (1 to 3) to your current '
 158 action_confidence':

158 - 1: Low confidence; contradictory or weak alignment across
 159 features.

159 - 2: Moderate confidence; partial alignment with moderate
 160 evidence.

160 - 3: High confidence; strong convergence across key features.

161 - Risk-profile adjustment:

162 - If 'risk_profile' is "High-Risk", and your trade is high risk
 163 high reward, consider increasing 'action_confidence' by +1, capped at 3.

163 - If 'risk_profile' is "Low-Risk" and downside risk is high,
 164 consider decreasing 'action_confidence' by -1, floored at 1.

164 - Feature Attribution:

```
165         - Rank the importance of each major feature used in your current
166         rationale using a Likert scale (1 to 3):
167             - 1: Minimal contribution; not required for the decision.
168             - 2: Moderate contribution; relevant but not critical.
169             - 3: High contribution; pivotal to the trading decision.
170         - If 'news_factors' are used, rank the top three with the same
171         Likert scale to reflect their impact on decision-making.
172 Output:
173 action: Str. LONG or SHORT.
174 action_confidence: int. Likert scale (1-3) confidence in the proposed 'action
175 ', adjusted based on prior strategy outcome if 'Last_Strategy_Used_Data' is
176 available.
177 explanation: >
178 A concise rationale (max 350 words) justifying the proposed 'action'.
179 Include:
180     - The top 5 weighted features used in the decision, each labeled with its
181     Likert importance (1 to 3).
182     (e.g., "Stock_Data.Price.Close, Weight 3, Technical_Analysis.RSI.Value,
183     Weight 1, Options_Data.ATM_Skew, Weight 2")
184     - If 'news_factors' were present, list the top 3 news signals with their
185     weights.
186     (e.g., "Earnings Call with positive analyst expectations, Weight 3.")
187     - A reflective assessment of 'Last_Strategy_Used_Data', including:
188         - Whether the past 'action' was successful and was it maintained given
189         prior 'Rationale'.
190         - Whether prior confidence was justified and how this influenced the
191         current decision.
```

C.9 | News Factors Prompt

```

1 User_Context:
2   Date: {date}
3   Ticker: {ticker}
4   Company_Name: {company_name}
5   Monthly_News_Articles_List: |
6     "{articles_list}"
7
8 System_Context:
9   Persona: Financial Market Analyst
10  Instructions: |
11  Extract the 'Top 5' news factors influencing stock price movements from the
12  'Monthly_News_Articles_List'. Follow these steps:
13    1. Rank the news by relevance to stock price movements:
14      - Prioritize news related to significant financial or market impacts (e
15      .g., acquisitions, partnerships, guidance revisions).
16      - Weigh industry trends, macroeconomic influences, and analyst ratings
17      based on their expected effect on the company's valuation.
18      - News with broad or long-term implications ranks higher.
19    2. Identify and anonymize only the target 'Company_Name' or 'Ticker', its
20    products, and employees. Do not anonymize other companies in the article:
21      - Replace 'Company_Name' with "THE COMPANY"
22      - Replace its products with generic terms such as "flagship product" or
23      "new software"
24      - Replace its employees with terms such as "THE CEO" or "a senior
25      executive"
26    3. Summarize content into key factors and corporate events affecting stock
27    prices, using concise language and causal relationships.
28
29  Examples of factors influencing stock prices include:
30    - Strategic partnerships or competitor activity.
31    - Industry trends or macroeconomic influences.
32    - Product launches or market expansions.
33    - Analyst ratings, significant stock price moves, or expectations.
34    - Corporate events: guidance revisions, acquisitions, contracts, splits,
35    repurchases, dividends.
36
37  Example:
38  'A major tech company partners with a leading automotive firm for EV battery
39  innovation. Analysts predict this could boost revenues significantly.'
40
41    - Ranked Factors:
42      1. Strategic partnership in EV battery technology expected to increase
43      revenue.
44      2. Positive sentiment driven by projected long-term gains.
45      3. Growing demand for EV technology anticipated to support future
46      earnings.
47
48 Deliver:
49   Factors: str array
50   - Extracted and ranked factors affecting the company from '
51   Monthly_News_Articles_List', with each factor summarized in no more than 120

```

words.

Data

D.1 | Replicated Benchmark Metrics

This appendix reports the replicated benchmark metrics for the assets used in Théate and Ernst (2021). We include mean SR, and MDD, each averaged across 25 runs with standard deviation σ .

For the SR, we compute a two-sided one-sample t -test to determine whether the mean of our replicated SR distribution significantly differs from the corresponding value reported in the original paper. The null hypothesis H_0 assumes that the replicated strategy has the same mean SR as the benchmark: $H_0 : \mu_{\text{SR}} = \text{SR}_{\text{paper}}$. The alternative hypothesis H_1 states that the replicated SR is significantly different: $H_1 : \mu_{\text{SR}} \neq \text{SR}_{\text{paper}}$. For metrics other than SR, p -values are omitted.

Bolded SR values indicate statistically significant differences ($p < 0.05$) from the original benchmark in table D.1. From the subset experimented on, only GOOGL failed to replicate the results.

D.2 | Data Availability

The equities, macroeconomic, and options datasets used in this thesis combine publicly licensed sources (FRED macroeconomic indicators, SEC EDGAR filings) with commercial subscriptions (Interactive Brokers historical market data, iVolatility implied volatility surfaces, and Alpaca news feeds). Public datasets are freely accessible under their respective data provider terms; commercial feeds require active licenses through the vendor portals. Derived features, prompts, and experiment logs that exclude proprietary price series are archived with the supervisory team and can be shared upon request via the author's supervisor, subject to the original licensing agreements.

Instrument	Paper SR	SR ($\pm\sigma$) [<i>p</i> -value]	MDD ($\pm\sigma$)
AB InBev	0.187	1.21 (0.30) [0.00]	0.18 (0.08)
Alibaba	0.021	0.06 (0.02) [0.00]	0.09 (0.01)
Amazon	0.419	0.39 (0.45) [0.85]	0.30 (0.09)
Apple	1.424	1.19 (0.55) [0.22]	0.29 (0.09)
Baidu	0.080	0.20 (0.17) [0.00]	0.36 (0.09)
CCB	0.202	0.33 (0.25) [0.04]	0.24 (0.14)
Coca Cola	1.068	1.07 (0.53) [0.50]	0.25 (0.04)
Dow Jones	0.684	0.70 (0.30) [0.91]	0.25 (0.05)
ExxonMobil	0.098	0.10 (0.35) [0.91]	0.34 (0.08)
FTSE 100	0.103	0.50 (0.23) [0.00]	0.31 (0.08)
Google	0.227	-0.54 (0.59) [0.00]	0.43 (0.13)
HSBC	0.011	0.38 (0.17) [0.00]	0.29 (0.05)
JPMorgan Chase	0.722	0.72 (0.31) [0.98]	0.26 (0.06)
Kirin	0.852	0.85 (0.42) [0.99]	0.39 (0.07)
Meta	0.151	0.63 (0.61) [0.01]	0.45 (0.27)
Microsoft	0.987	0.70 (1.00) [0.38]	0.28 (0.16)
NASDAQ 100	0.845	0.85 (0.35) [1.00]	0.16 (0.05)
Nikkei 225	0.019	0.26 (0.29) [0.02]	0.29 (0.07)
Nokia	-0.094	0.07 (0.24) [0.00]	0.57 (0.15)
PetroChina	0.156	0.22 (0.29) [0.29]	0.67 (0.00)
Philips	0.675	1.40 (0.50) [0.00]	0.25 (0.03)
S&P 500	0.834	0.83 (0.25) [1.00]	0.14 (0.04)
Shell	0.425	0.42 (0.37) [0.95]	0.51 (0.05)
Siemens	0.426	0.39 (0.23) [0.43]	0.26 (0.12)
Sony	0.424	0.42 (0.36) [0.97]	0.16 (0.04)
Tesla	0.621	0.48 (0.41) [0.29]	0.52 (0.09)
Tencent	-0.198	-0.19 (0.33) [0.98]	0.10 (0.09)
Toyota	0.304	0.36 (0.27) [0.37]	0.45 (0.10)
Volkswagen	0.216	0.45 (0.18) [0.00]	0.48 (0.09)

Table D.1: Replication Metrics for Théate and Ernst (2021)

Replication and Reproducibility

This appendix documents the experimental environment, configuration artefacts, and preprocessing routines required to reproduce the results reported in Chapter 3.2.2–4.3. The material below complements the methodological description in Chapter 3.

E.1 | Software Environment

Table E.1: Core software versions used in all experiments

Component	Version
Operating System	Ubuntu 22.04.4 LTS (5.15.0-117-generic kernel)
Python	3.10.14 (Conda <code>quant_drl</code> environment)
CUDA Toolkit	12.4.1 (conda-forge build, matching NVIDIA 550.xx drivers)
PyTorch	2.4.1 (pip, CUDA 12.1 wheels)
TensorFlow	2.15.0 (CUDA 12.0 build)
Transformers	4.41.2
Pandas	2.2.3
NumPy	1.26.4
SciPy	1.15.1
scikit-learn	1.4.0
TA-Lib	0.4.32
statsmodels	0.14.1
Matplotlib	3.8.2
TF-Agents	0.19.0 (pip, Experiment 2 baselines)
OpenAI Python SDK	1.51.2
Biber	2.19

Source code, Conda environment files, and experiment notebooks are archived with the supervisory team and are available on request.

E.2 | Hardware Configuration

- CPU: AMD Ryzen 9 5900HS (8 cores / 16 threads, base 3.0 GHz)
- GPU: NVIDIA GeForce RTX 3050 Laptop GPU (4 GB VRAM)
- RAM: 32 GB DDR4 3200 MHz
- Storage: 1 TB NVMe SSD with 550 MB/s sustained write throughput

Chapter 3.2.3.4 already reports aggregate run times. In practice, a single equity run (25 replications × 50 episodes) required ~3 hours wall-clock time using the above hardware.

E.3 | Random Seeds and Replications

Deterministic seeding is enforced across PyTorch, NumPy, and Python's `random` module. The following base seeds drive each replication batch:

- Prompt-tuning loop (writer–judge memory): base seed 45
- Experiment 1 back-testing windows: base seed 45
- Experiment 2a (high-risk DDQN training): base seed 45
- Experiment 2b (low-risk DDQN training): base seed 45

Each seed tuple sets the Python, NumPy, and PyTorch generators before environment creation, ensuring reproducible replay buffers and epsilon-greedy action schedules.

E.4 | Data Processing

- **Chronological slicing:** Data are sorted by timestamp and partitioned into 2012–2018 (train), 2018–2020 (test) folds with no embargo. This follows the same scheme as Théate and Ernst (2021).
- **Feature engineering:** All technical indicators, fundamental ratios, and macroeconomic features are computed as per Section 3.2.2.1. Rolling windows for indicators are aligned to the last trading day of each month to prevent look-ahead bias.

- **Scaling:** Market features use rolling min–max scaling over 252 trading days; fundamental and macro indicators use z-scores computed on the training fold and applied out-of-sample.
- **Missing values:** Forward-fill within each security; all data is verified to contain no missing values.
- **Entity and Date Anonymization:** Entity and date placeholders are anonymized as described in Section 3.2.2.1, ensuring prompts do not expose post-cutoff knowledge.

E.5 | Prompt and Model Configuration

- Language model: OpenAI GPT-4o mini, API snapshot 2024-05-13, temperature 0.7, top-k 3, frequency penalty 1.0, presence penalty 0.25.
- Strategy cadence: One prompt call per instrument per calendar trading month (20 days); completions cached and reused across RL episodes.
- Reward shaping: Replicated Théate and Ernst (2021) reward with zero risk-free rate; SR and MDD computed on daily returns.
- Logging: All completions, evaluator scores, and entropy/perplexity diagnostics stored as JSONL files for audit.

In reproducing the experiments, readers should confirm GPU drivers equal to or newer than NVIDIA 576.02 and verify compatibility of the OpenAI API snapshot used at inference time. If API updates deprecate the 2024-05-13 snapshot, the cached prompts maintained alongside this appendix should be replayed to approximate the originally reported strategies.

Expert Reviews

For the expert reviews of the system's strategies, human subjects with relevant financial experience were asked to complete a questionnaire. The questionnaire was distributed via Google Forms⁹.

The attached booklet below provides the full questionnaire and summarized responses, the demographic summary is available in Table F.1.

Characteristic	Distribution (%)
Gender (Male / Female)	90 / 10
Age 25–34 / 35–44 / 45–54 / 55+	30 / 50 / 10 / 10
Professional / Retail Background	80 / 20

Table F.1: Expert Review Participant Demographics (Aggregated Summary)

⁹https://docs.google.com/forms/d/e/1FAIpQLSez3b7huMyt_HHuoKVyCI0yZfTCituDThFVHPBj2Won-xHxQ/viewform

Expert Evaluation of LLM-Generated Trade Strategies

Instructions for the review below:

1. **Step 1: Problem Definition:** The first task is to analyze the given information and determine the likely optimal trading action for the next month: *LONG* (buy) or *SHORT* (sell) with a brief justification (one or two sentences) for your chosen position.
2. **Step 2: Expert Scoring:** Evaluate the LLM-generated trade strategy by analyzing its decision, *rationale* and *selection of key influencing factors*, you will use this [rubric](#).
3. **Step 3: Expert Exemplars:** Provide specific feedback on errors in the form of corrective explanations . You can skip this if the model was perfect, or you're unsure what to correct.

Consent: By proceeding with this survey, you consent to the use of your responses in this research study. Your email will only be used to limit one set of answers per reviewer and to share results. It will be deleted at the end of the research.

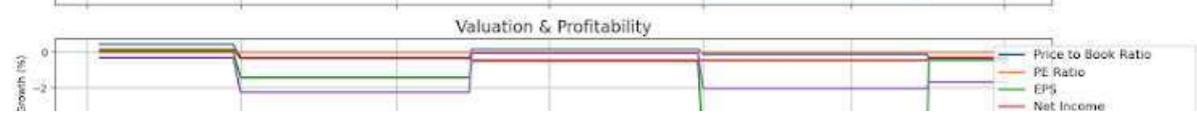
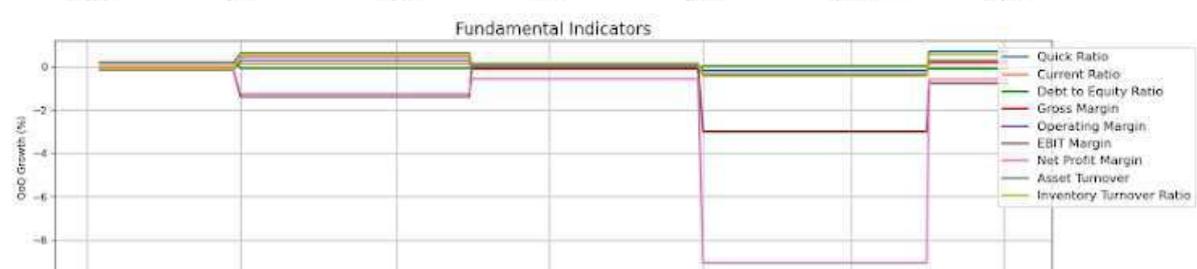
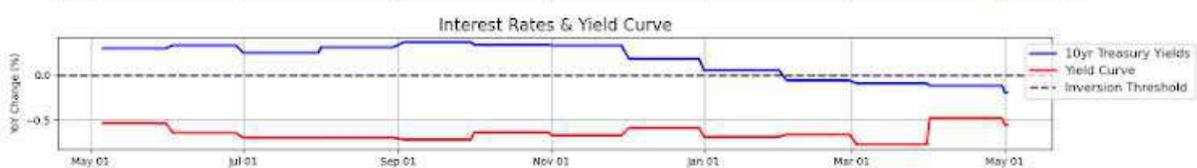
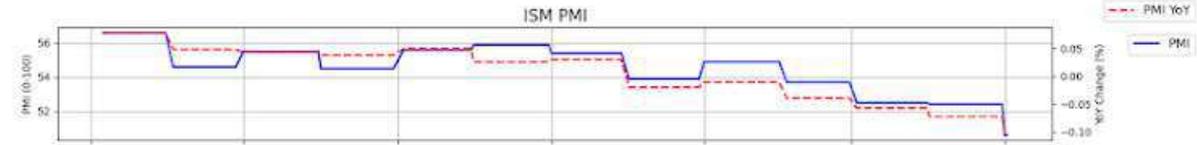
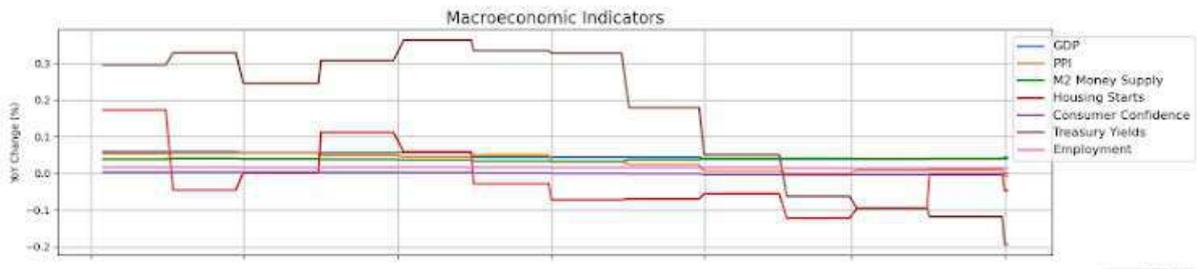
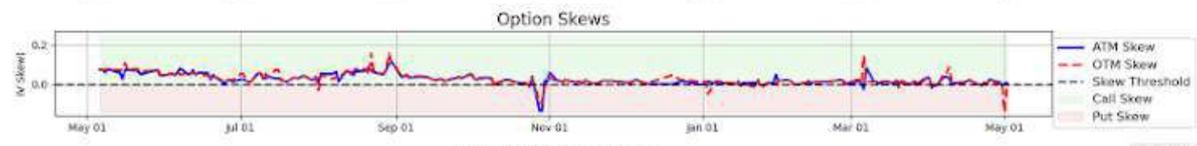
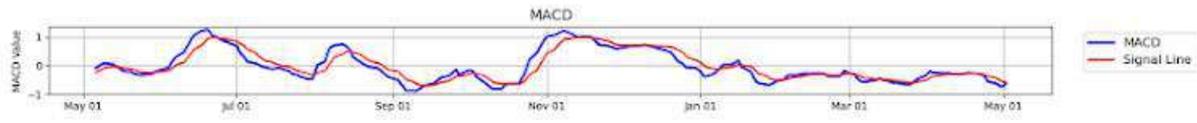
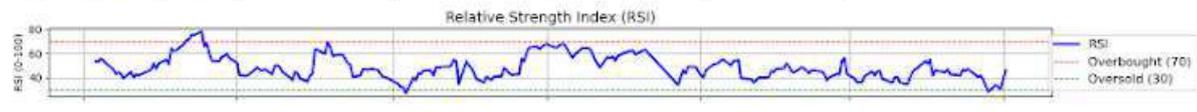
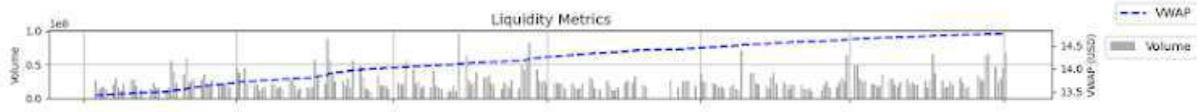
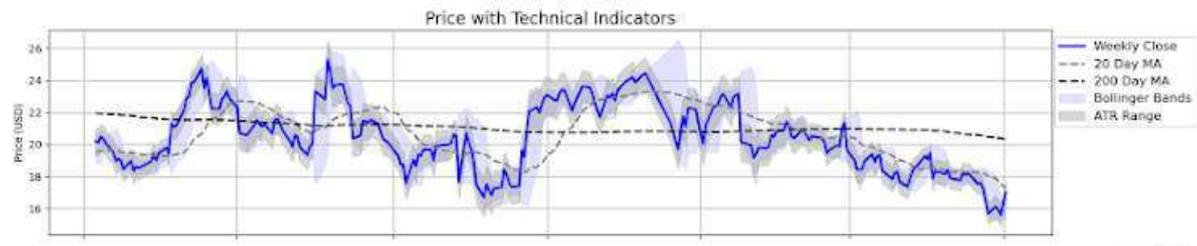
* Indicates required question

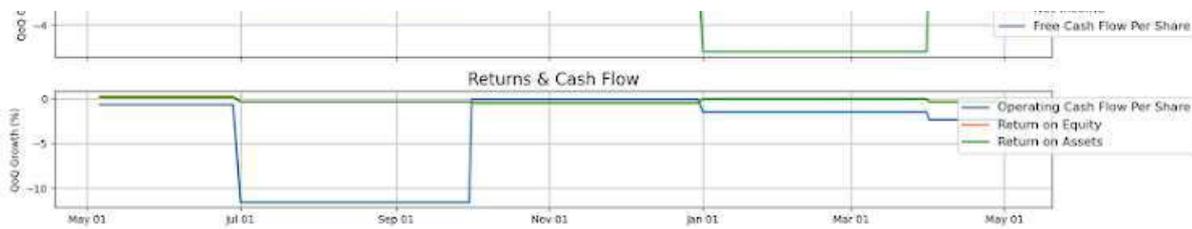
1. Email *

2. **Optional:** If you wish to be acknowledged in the research publications "*Language Model Guided Reinforcement Learning in Quantitative Trading*", please provide your name, surname, and affiliation below.

Review of LLM [Strategy 1](#)

Historic 1 Year Data





Historic 1 Month News

Board Restructuring & Corporate Governance Changes THE COMPANY announced plans to streamline its board by reducing the number of directors and changing voting procedures. Some investors saw this as a positive step toward improved corporate governance, while others expressed concerns about reducing independent oversight. The impact on THE COMPANY’s stock was initially neutral to slightly negative.

Vehicle Fire Incident in China A viral video showed a parked vehicle from THE COMPANY spontaneously combusting in China. This incident triggered strong reactions from Chinese consumers and raised concerns about vehicle safety, particularly in a key market where THE COMPANY is trying to expand. The news contributed to a stock decline and added pressure on its reputation.

Autonomous Driving & Robotaxi Claims THE CEO made bold claims about launching fully autonomous taxis by next year, stating that owning any other vehicle would soon be "financially insane." While some analysts acknowledged THE COMPANY’s leadership in electric vehicles, others viewed the announcement as an attempt to distract from operational challenges. Investors were skeptical, and the stock reaction remained muted.

Disappointing Earnings Report THE COMPANY reported worse-than-expected financial results, missing both revenue and earnings estimates. Vehicle deliveries declined, cash reserves fell, and the outlook for profitability weakened. Analysts downgraded the stock, citing concerns over demand, production efficiency, and capital requirements. The stock dropped following the earnings release.

Capital Raise Speculation Following the earnings report, analysts predicted that THE COMPANY would need to raise additional funds to sustain operations and growth. THE CEO initially resisted the idea but later acknowledged its potential necessity. This uncertainty around financing created volatility in the stock and raised concerns about long-term financial stability.

3. Guess the Trade Action for the next month: *

Mark only one oval.

LONG

SHORT

4. Provide your rationale *

LLMs' Strategy

The traded instrument was TSLA, with the trading period set between 2019-05-05 and 2019-06-02.

LLMs' decision: *SHORT*

Rationale: The analysis suggests a SHORT position due to several concerning factors. Firstly, the company's fundamental metrics are weak, with negative EPS YoY (-46.43%), Free Cash Flow YoY (-183.06%), and Net Income YoY (-75.37%). These figures indicate declining profitability and cash flow issues, which undermine confidence in future earnings growth.

Secondly, the technical indicators reveal bearish signals: the stock is currently trading below its 50MA (19.52) and 200MA (20.89), suggesting that momentum is shifting downward. The MACD shows a bearish strength (-0.286) while RSI sits at 51, indicating neutral but potential weakness ahead.

Furthermore, news factors highlight significant risks: analysts have downgraded the stock rating from Outperform to Peer Perform due to concerns over demand for its flagship product—an indication of weakening consumer interest amidst increasing competition in the electric vehicle market.

Top weighted features influencing this decision:

Stock_Data.Price.Close (19.06)

Technical_Analysis.MACD.Value (-0.286)

Fundamental_Data.Growth_Metrics.EPS_YoY (-46%)

Fundamental_Data.Growth_Metrics.Free_Cash_Flow_YoY (-183%)

Options_Data.Put_IV_Skews.Otm_Skew (1.05)

Key news impacting this strategy:

Analysts' downgrade on performance - Market Impact Score: 5/5

CEO's share purchase showing confidence amid uncertainties - Market Impact Score: 4/5

Fatal crash raising safety concerns leading to regulatory scrutiny - Market Impact Score: 4/5

LLM Decision Insights

Long Probability 0.1

Short Probability 0.9

Perplexity 1.86614

Total Tokens 19627

Total Cost (USD) \$0.003212

Performance Metrics

Sharpe Ratio (Annualized SR) 7.35114

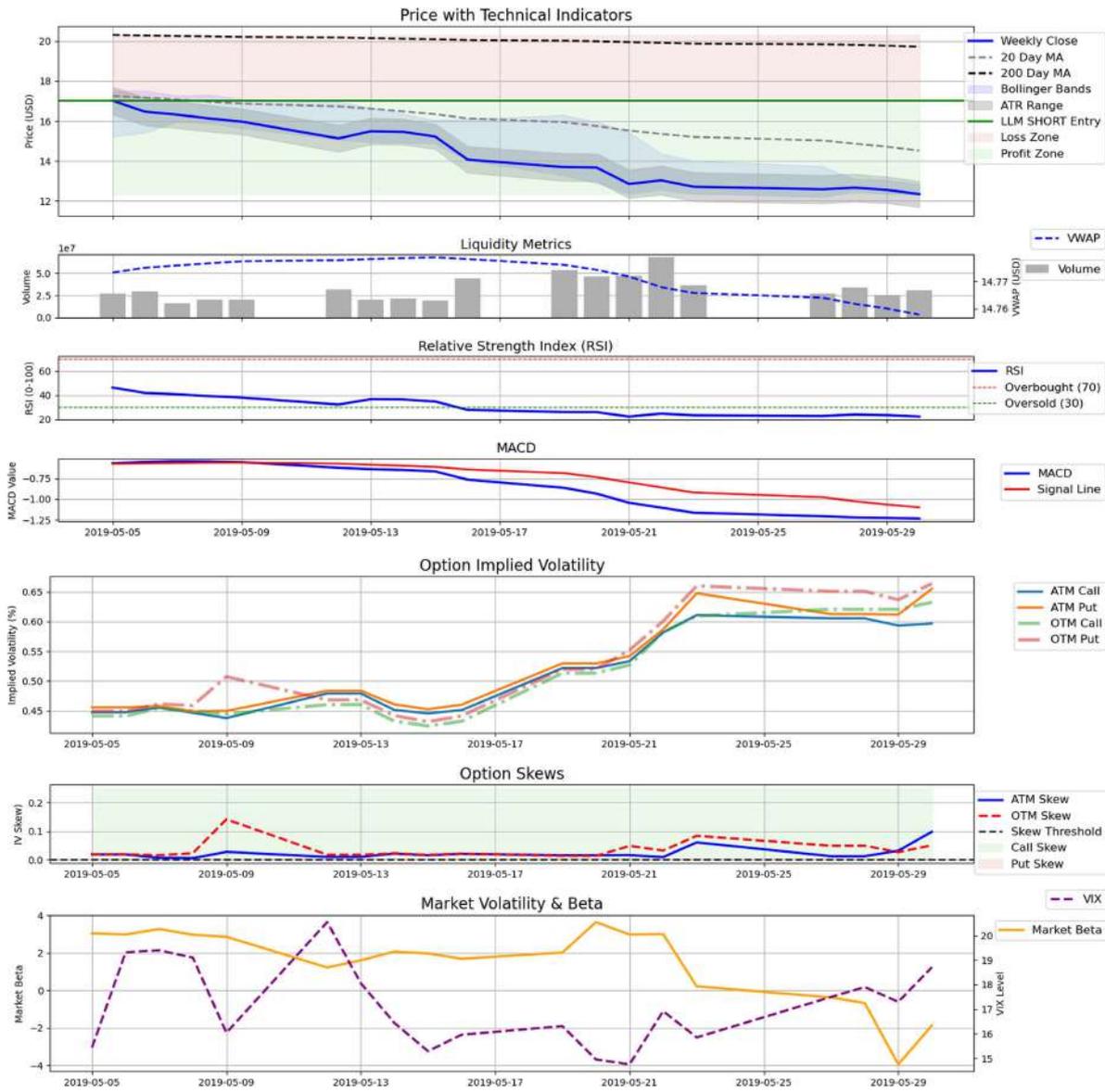
Portfolio Turnover (PTR) 0

Maximum Drawdown (MDD) 0.0459915

Mean Drawdown Duration (MDDur) 2

Cumulative Returns 0.294539

Following Month Data



5. Judge the LLMs' [Action](#) using this [Rubric](#). *

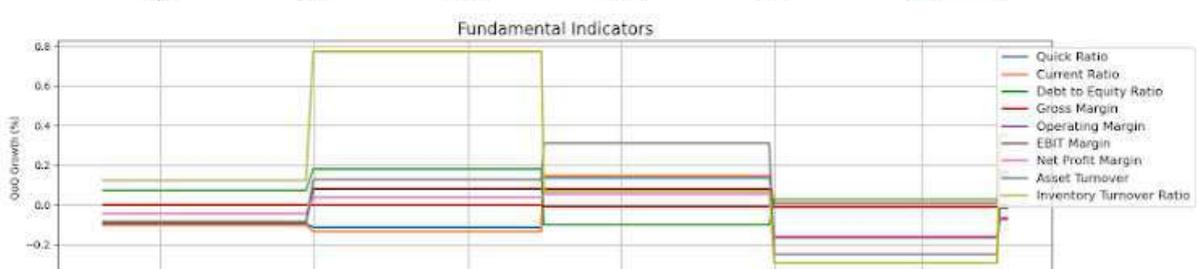
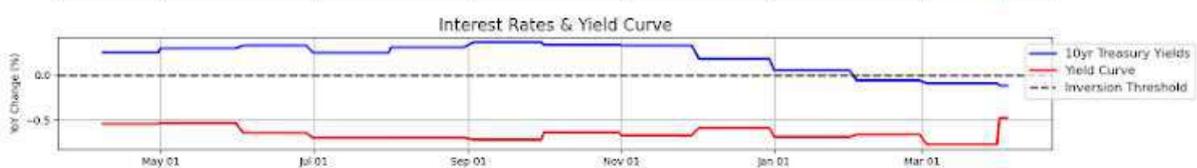
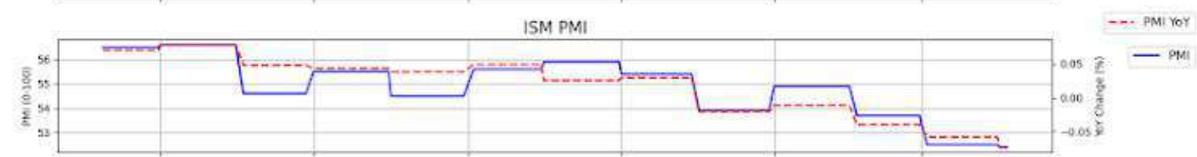
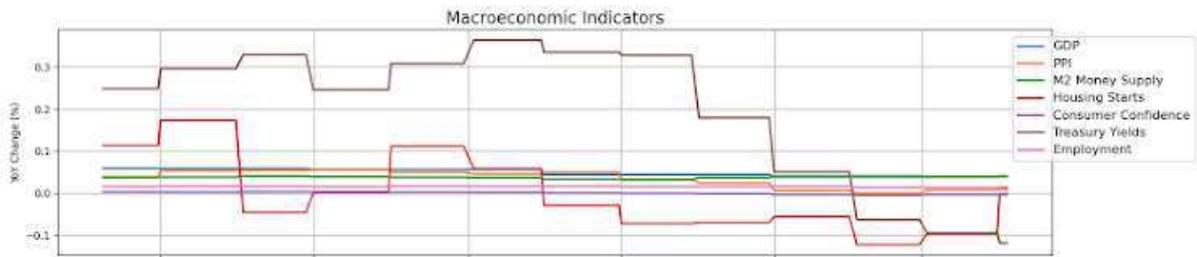
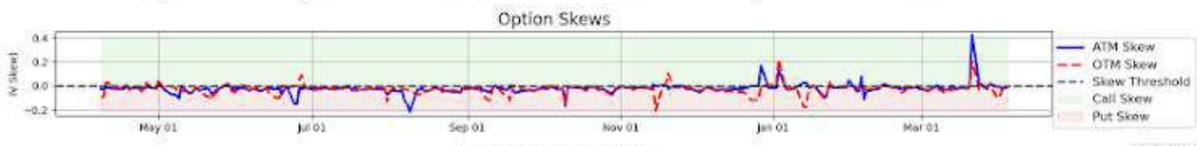
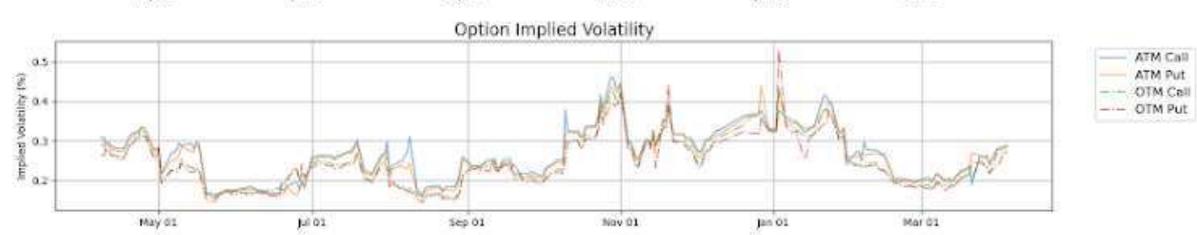
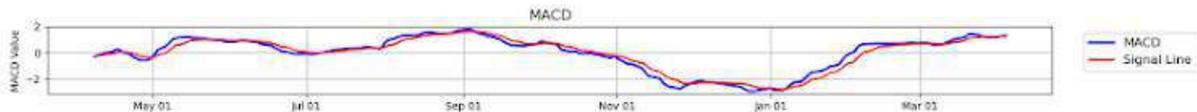
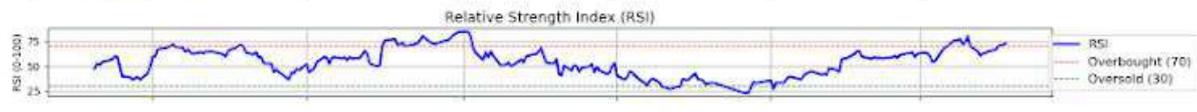
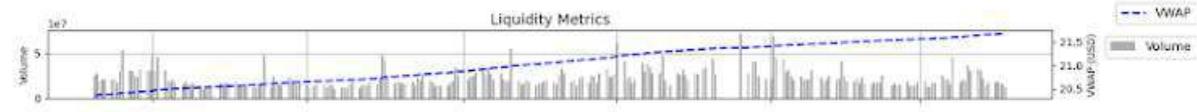
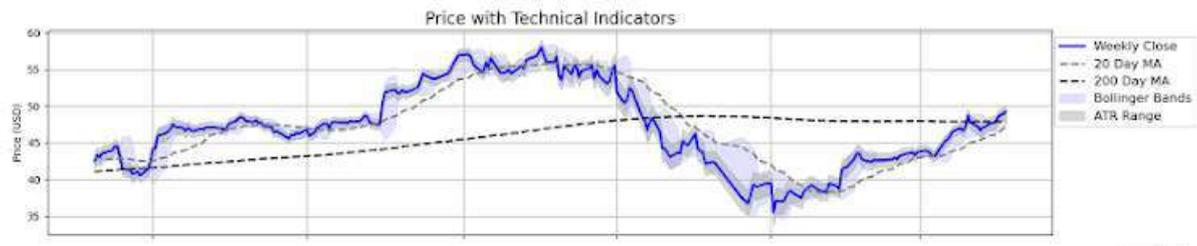
Mark only one oval per row.

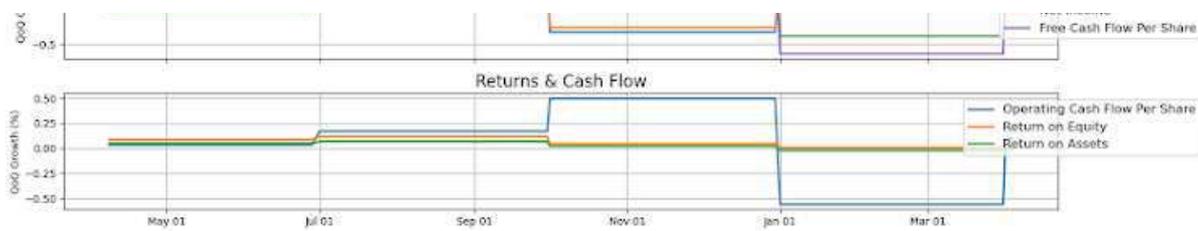
	Score -1 (Poor)	Score 0 (Average)	Score 1 (Good)
Rationale	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fidelity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Safety	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

6. Provide your [exampler](#) (if applicable).

Review of LLM Strategy 2

Historic 1 Year Data





Historic 1 Month News

Increased Regulatory Scrutiny on Tech Companies Amid growing concerns over data privacy and antitrust regulations, government agencies are considering imposing stricter oversight on major tech firms, including THE COMPANY. Analysts speculate that potential legal battles or regulatory fines could impact THE COMPANY's ability to innovate freely and expand its services sector.

Supply Chain Disruptions Affecting Production Recent reports highlight supply chain constraints affecting key components in THE COMPANY's flagship product. Delays in procurement and rising material costs have forced THE COMPANY to adjust production schedules, raising concerns about meeting consumer demand in the upcoming quarters.

Expansion of THE COMPANY's Services Sector THE COMPANY is accelerating its push into subscription-based services, introducing new offerings aimed at diversifying revenue streams beyond hardware sales. Analysts view this as a strategic move to reduce dependency on product cycles and strengthen long-term financial stability.

Legal Dispute Over Intellectual Property A recent legal battle involving THE COMPANY and a competitor over key patents in software and hardware integration has escalated. While THE COMPANY asserts its technology leadership, potential litigation outcomes could impact future product developments and licensing agreements.

Market Reaction to Economic Indicators Recent economic data, including consumer spending trends and inflation reports, have influenced investor sentiment toward THE COMPANY. Market analysts are closely watching how macroeconomic conditions will affect THE COMPANY's pricing strategy and global sales performance.

7. Guess the Trade Action for next month: *

Mark only one oval.

LONG

SHORT

8. Provide your rationale *

LLMs' Strategy

The traded instrument was AAPL, with the trading period set between 2019-04-07 and 2019-05-05.

LLMs' decision: *LONG*

Rationale: The analysis indicates a bullish strategy for the next month based on several compelling factors. First, the stock's current close price of 44.05 is above both its 20MA (42.11) and 50MA (42.86), suggesting upward momentum (Feature: Stock_Data.Price.Close). Second, with an RSI value of 58.26, it indicates that the stock is not overbought yet and may have room to run higher before reaching overbought conditions (Feature: Technical_Analysis.RSI.Value). Third, despite recent volatility indicated by a historical volatility of approximately 27.78%, the implied volatility at around 22.63% suggests that options are priced favorably for potential upside moves (Feature: Implied_Volatility.IV_Close). Fourth, fundamental growth metrics such as EPS YoY at approximately 28% signal strong earnings potential which supports price appreciation (Feature: Fundamental_Data.Growth_Metrics.EPS_YoY). Lastly, news analysis reveals positive sentiment from market analysts regarding improved production capabilities; this aligns with pricing trends and reinforces confidence in future profitability due to reduced dependence on external suppliers.

Top weighted features influencing this decision:

Stock_Data.Price.Close (44.05)

Technical_Analysis.RSI.Value (58.26)

Implied_Volatility.IV_Close (22.63%)

Fundamental_Data.Growth_Metrics.EPS_YoY (28%)

Historical_Volatility (27.78%)

Key news impacting this strategy:

Market analysts upgraded ratings on THE COMPANY due to improved production capabilities - Market Impact Score: 4/5

The strategic shift to self-manufactured chips potentially enhancing profit margins - Market Impact Score: 4/5

Mixed sentiments about production targets indicating cautious optimism but manageable risk - Market Impact Score: 3/5

LLM Decision Insights

Long Probability 0.65

Short Probability 0.35

Perplexity 9.14316e+15

Total Tokens 38845

Total Cost (USD) \$0.006040Following Month Data

Performance Metrics

Sharpe Ratio (Annualized SR) 1.89417

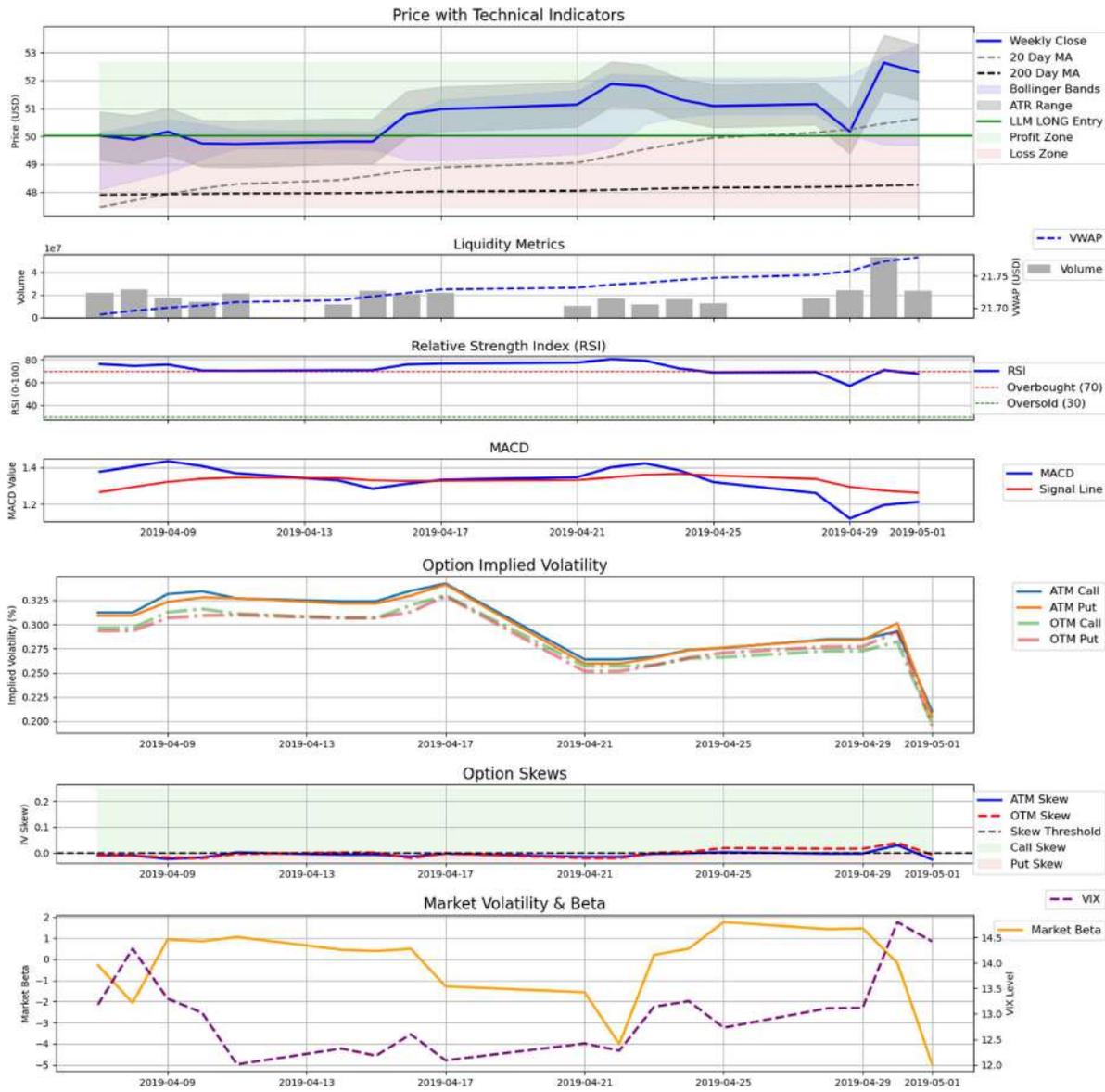
Portfolio Turnover (PTR) 0

Maximum Drawdown (MDD) 0.0893178

Mean Drawdown Duration (MDDur) 5

Cumulative Returns 0.0484561

Following Month Data



9. Judge the LLMs' [Action](#) using this [Rubric](#). *

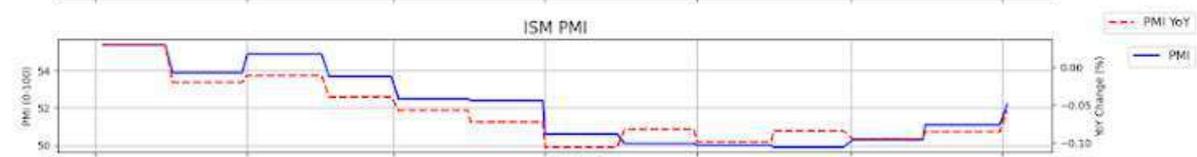
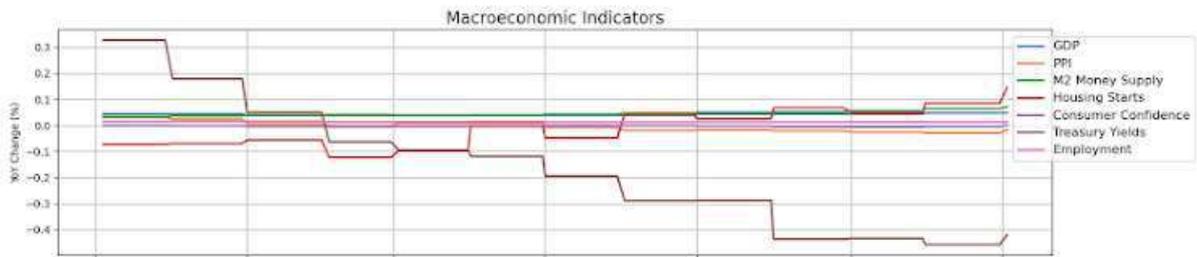
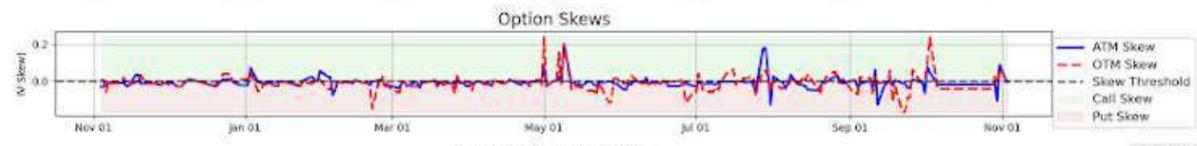
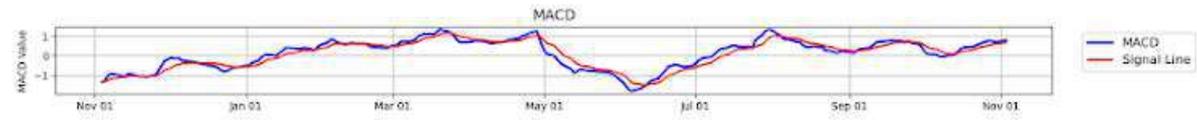
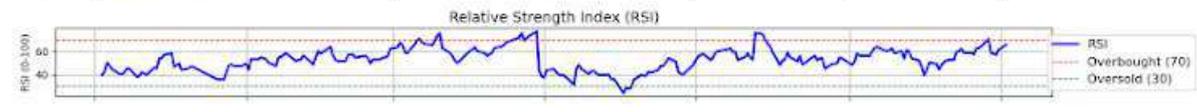
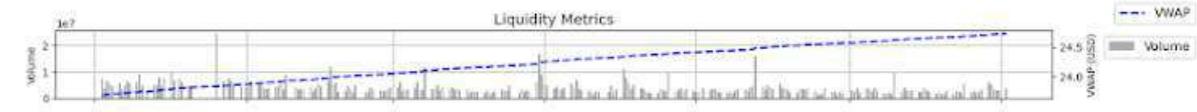
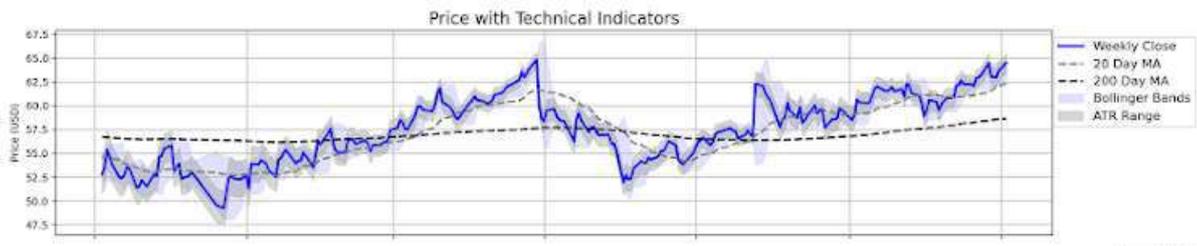
Mark only one oval per row.

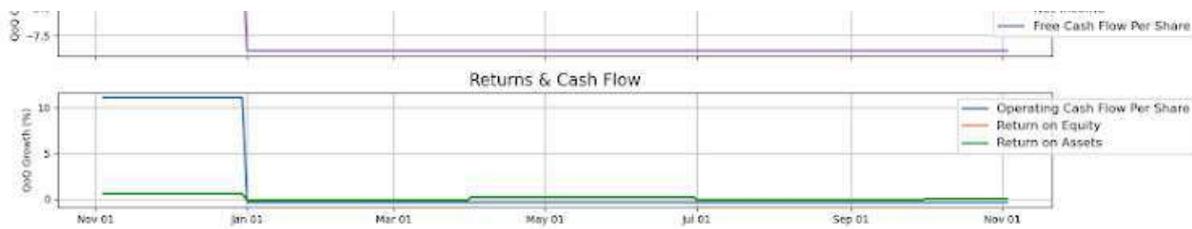
	Score -1 (Poor)	Score 0 (Average)	Score 1 (Good)
Rationale	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fidelity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Safety	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

10. Provide your [exampler](#) (if applicable).

Review of LLM Strategy 3

Historic 1 Year Data





Historic 1 Month News

Board Investigation: THE COMPANY's board is investigating allegations of sexual harassment against senior executives, which could impact investor confidence and stock performance. The investigation follows a lawsuit claiming the board covered up misconduct, raising concerns about corporate governance and accountability.

Healthcare Data Partnership: THE COMPANY's partnership with a major hospital system to store and analyze patient data has raised privacy concerns among consumers. Although the initiative complies with federal laws, the lack of patient consent could lead to reputational damage and affect stock performance.

Competitive Pressure: A significant product launch by a competitor, including new software and mobile applications, could shift market dynamics and impact THE COMPANY's competitive position in the industry.

Product Performance: Analysts have noted a strong increase in downloads for THE COMPANY's flagship product, indicating positive user engagement and potential revenue growth, which could positively influence stock prices.

Stock Performance: THE COMPANY's stock reached a new 52-week high, reflecting positive market sentiment and investor confidence in its growth trajectory, potentially leading to further price appreciation

11. Guess the Trade Action for next month: *

Mark only one oval.

LONG

SHORT

12. Provide your rationale *

LLMs' Strategy

The traded instrument was GOOGL, with the trading period set between 2019-11-04 and 2019-12-02.

LLMs' decision: *LONG*

Rationale: The trading strategy is to take a long position on THE COMPANY based on the following analysis. The stock's closing price of 63.61 is above both the 20MA (62.09) and 50MA (61.1), indicating bullish momentum. This is supported by a positive MACD crossover and an RSI of 61.31, suggesting moderate upward movement without being overbought. Weekly returns have fluctuated but show signs of recovery with a recent uptick of 2.17%. With a high beta of 2.88, the stock is sensitive to market movements, making it suitable for aggressive trading under a high-risk profile.

Key news factors include analysts maintaining a positive outlook due to strategic moves in financial services aimed at monetizing their user base, a significant increase in downloads for its flagship product indicating strong user engagement and potential revenue growth, and ongoing investigations into allegations against senior executives posing reputational risks. Despite these risks, stronger fundamentals and positive market sentiment support leveraging the current bullish technical indicators.

Top weighted features influencing this decision:

Stock_Data.Price.Close (61.15)

Technical_Analysis.RSI.Value (39.67)

Macro_Indicators.GDP_QoQ (0.0099)

Macro_Indicators.PMI (51.1)

Implied_Volatility.IV_Close (-0.457)

Key news impacting this strategy:

Analysts predict significant rebounds in deliveries - Market Impact Score: 4/5

Strategic collaboration expected to enhance product offerings - Market Impact Score: 5/5

Analyst upgrades suggest bullish outlook - Market Impact Score: 4/5

LLM Decision Insights

Long Probability 0.75

Short Probability 0.25

Perplexity 5.36689e+29

Total Tokens 21817

Total Cost (USD) \$0.003487Following Month Data

Performance Metrics

Sharpe Ratio (Annualized SR) -0.299606

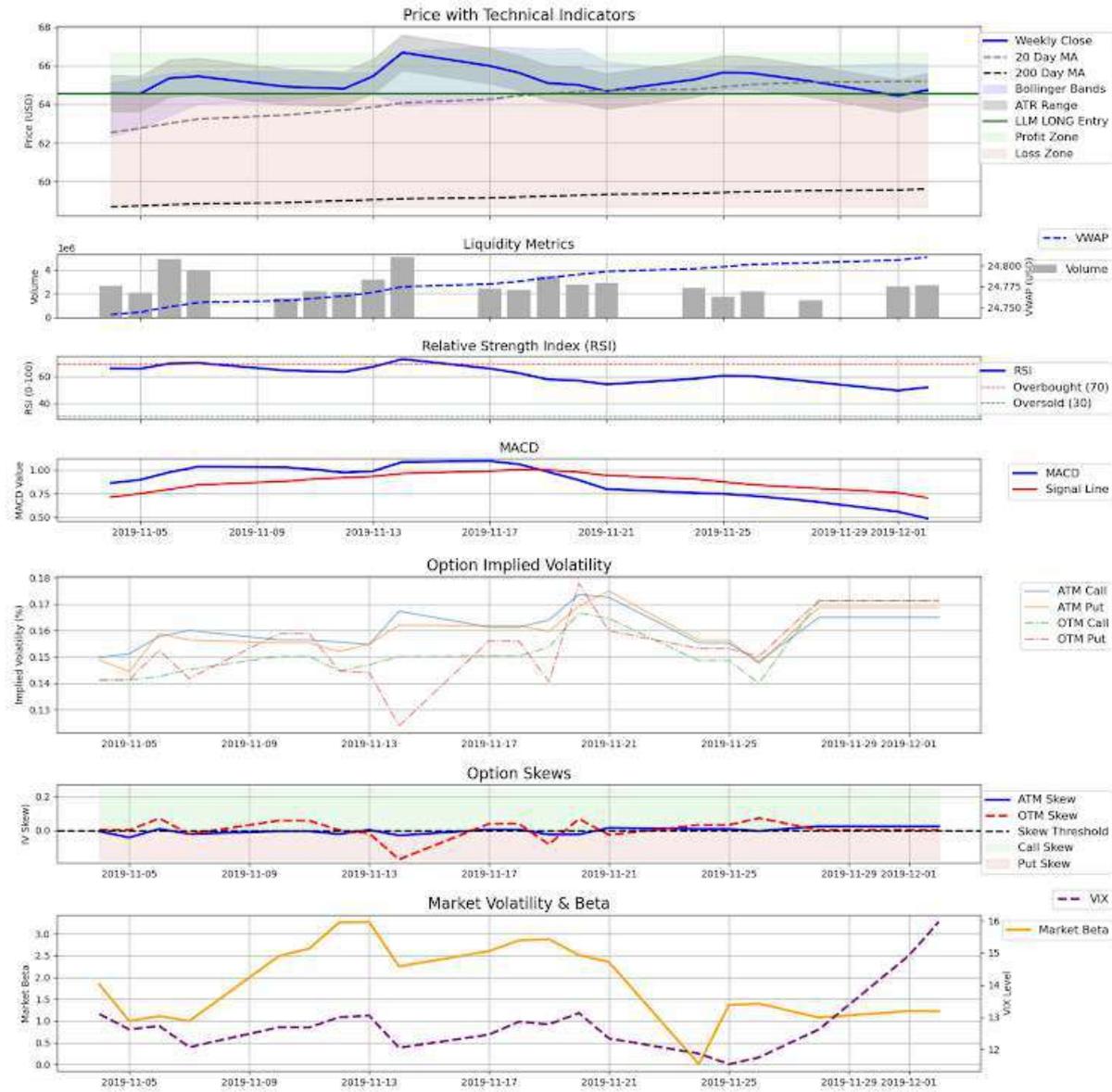
Portfolio Turnover (PTR) 0

Maximum Drawdown (MDD) 0.0335933

Mean Drawdown Duration (MDDur) 4.42857

Cumulative Returns -0.00201332

Following Month Data



13. Judge the LLMs' [Action](#) using this [Rubric](#). *

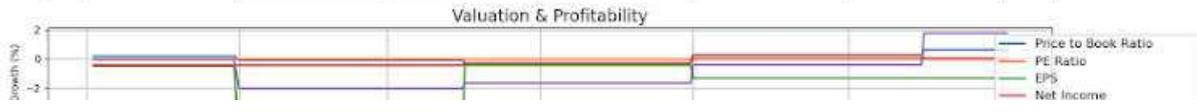
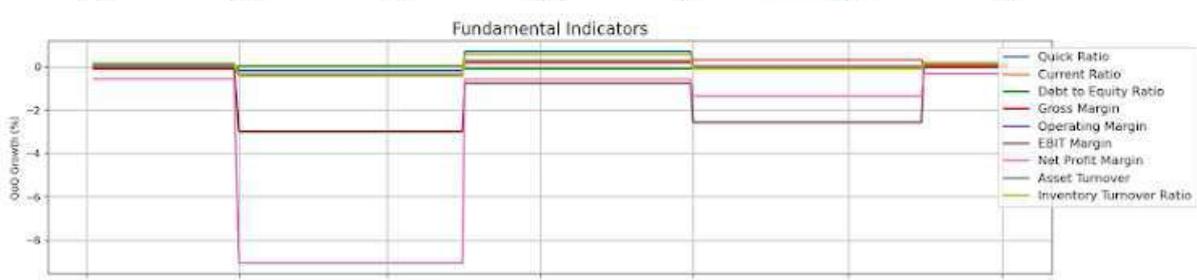
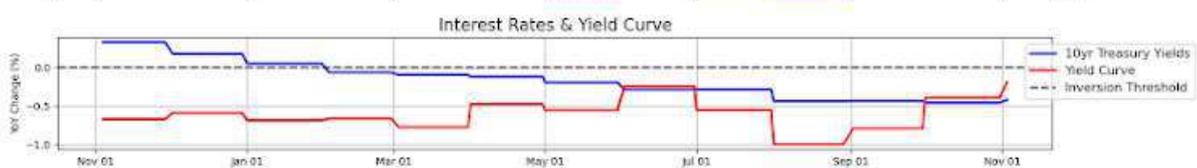
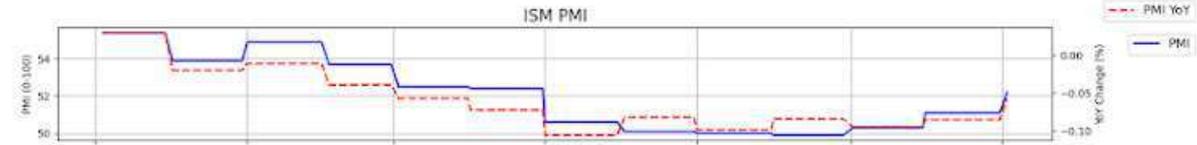
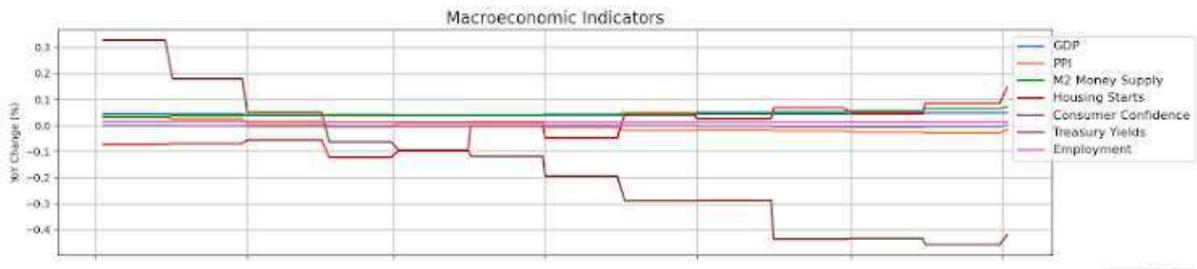
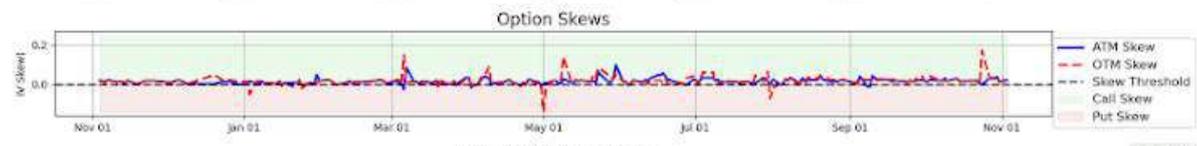
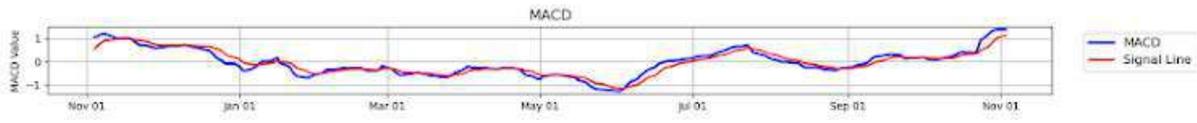
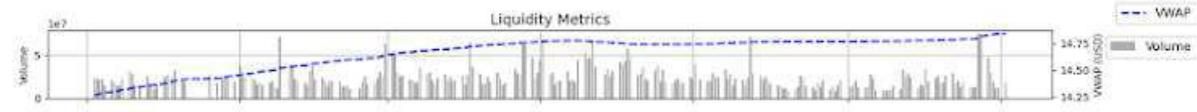
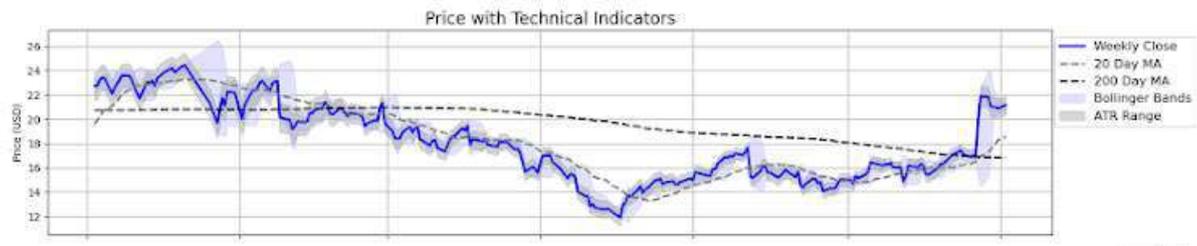
Mark only one oval per row.

	Score -1 (Poor)	Score 0 (Average)	Score 1 (Good)
Rationale	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fidelity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Safety	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

14. Provide your [exampler](#) (if applicable).

Review of LLM Strategy 4

Historic 1 Year Data





Historic 1 Month News

Earnings Report Surpasses Expectations THE COMPANY reported earnings significantly above market expectations, with stronger-than-anticipated revenue and profitability. Analysts had projected a loss, but THE COMPANY's financials demonstrated resilience, driven by improved operational efficiency and cost management.

Strategic Partnership to Enhance Product Offerings THE COMPANY announced a strategic collaboration with a major technology firm to enhance its next-generation product capabilities. This move is expected to improve market competitiveness and expand offerings, reinforcing its position in the industry.

Market Confidence Boosted by Analyst Upgrades Several analysts revised their ratings upward for THE COMPANY, citing strong future growth prospects and improving financial stability. Investor sentiment turned more positive following these upgrades, leading to increased stock momentum.

Macroeconomic Factors Favoring THE COMPANY Economic indicators, including GDP growth and declining treasury yields, have contributed to a favorable market environment for THE COMPANY. These factors have led to increased investor confidence in high-growth stocks, benefiting THE COMPANY's valuation.

Competitive Market Pressures Intensify A major competitor introduced a new product line, intensifying competition within the sector. While this development poses challenges, THE COMPANY's recent strategic initiatives and product advancements are expected to mitigate the competitive impact.

15. Guess the Trade Action for next month: *

Mark only one oval.

LONG

SHORT

16. Provide your rationale *

LLMs' Strategy

The traded instrument was TSLA, with the trading period set between 2019-11-04 and 2019-12-02.

LLMs' decision: *LONG*

Rationale: The decision to pursue a LONG strategy is supported by multiple factors indicating bullish momentum and favorable market conditions. Key features include:

Stock_Data.Price.Close: The current price of 16.21 is above the 20MA (15.96) and the 50MA (15.50), suggesting positive momentum.

Technical_Analysis.RSI.Value: With an RSI of 55, the stock indicates potential for further upside as it approaches overbought territory (>70).

Options_Data.Atm_IV_Skews: The ATM IV skew shows stability, which coupled with high implied volatility suggests readiness to capitalize on upward price movement.

Growth Metrics (EPS YoY): Although EPS YoY is negative at -16%, recent news suggests expected recovery due to new product demand and operational efficiencies from new facilities.

Macro_Data.SPX.Close_Slope: The SPX close slope indicates a slight bearish trend; however, VIX trends are showing reduced fear in the market, supporting risk-on behavior.

Key news impacting this strategy:

Positive delivery statistics reported by THE COMPANY indicate a rebound in sales - Market Impact Score: 5/5

Analysts raise price targets citing anticipated revenue growth - Market Impact Score: 4/5

THE COMPANY announces plans for a new Gigafactory in Berlin, enhancing production efficiency - Market Impact Score: 4/5

LLM Decision Insights

Long Probability 0.75

Short Probability 0.25

Perplexity 8.61386e+12

Total Tokens 21937

Total Cost (USD) \$0.003532

Performance Metrics

Sharpe Ratio (Annualized SR) 2.18892

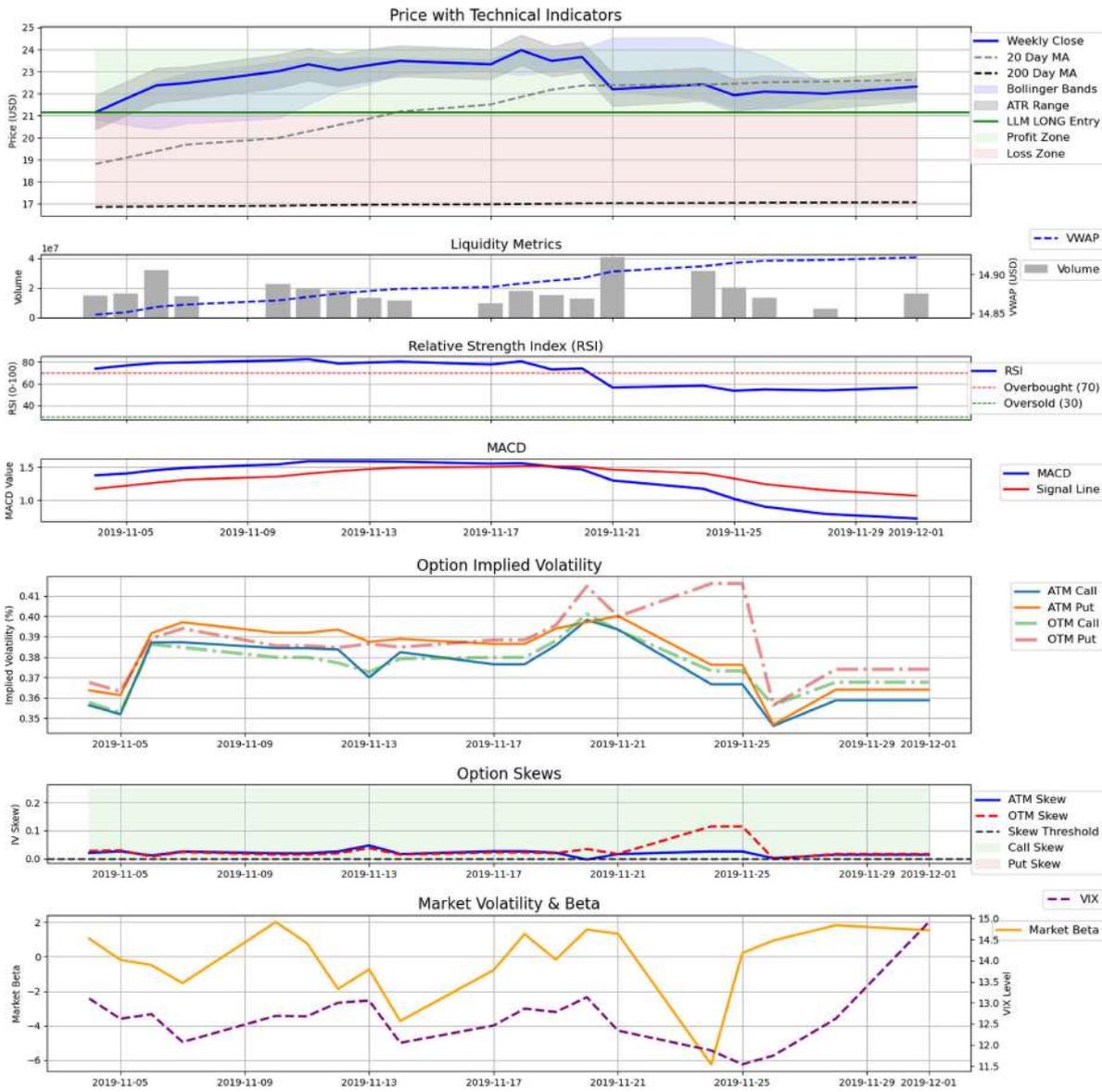
Portfolio Turnover (PTR) 0

Maximum Drawdown (MDD) 0.0851064

Mean Drawdown Duration (MDDur) 3.63636

Cumulative Returns 0.0553191

Following Month Data



17. Judge the LLMs' [Action](#) using this [Rubric](#). *

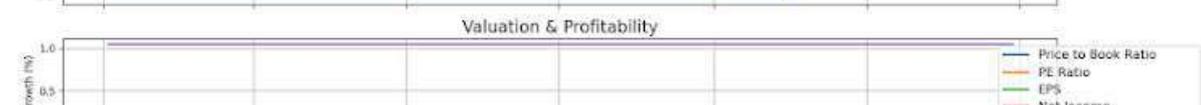
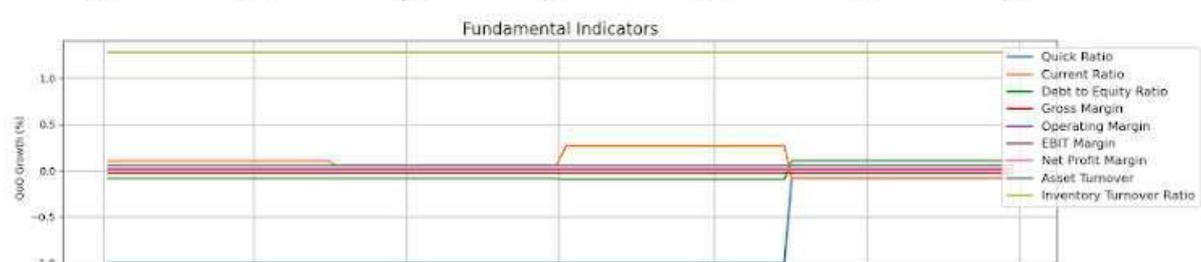
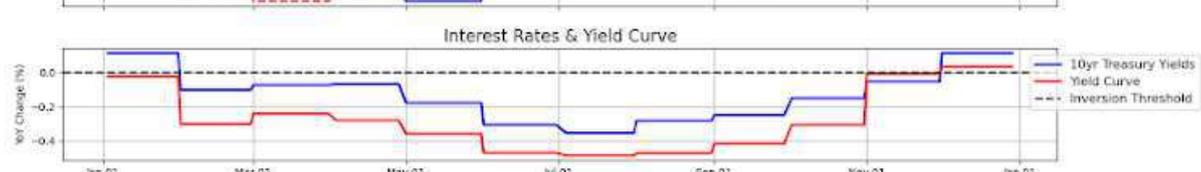
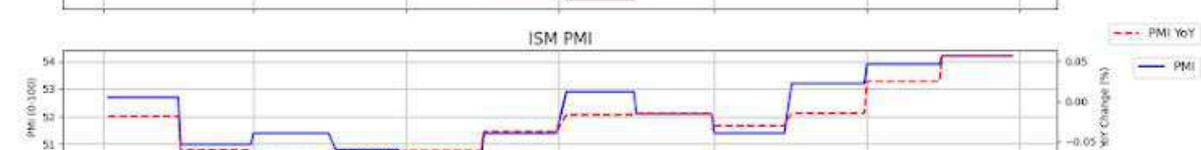
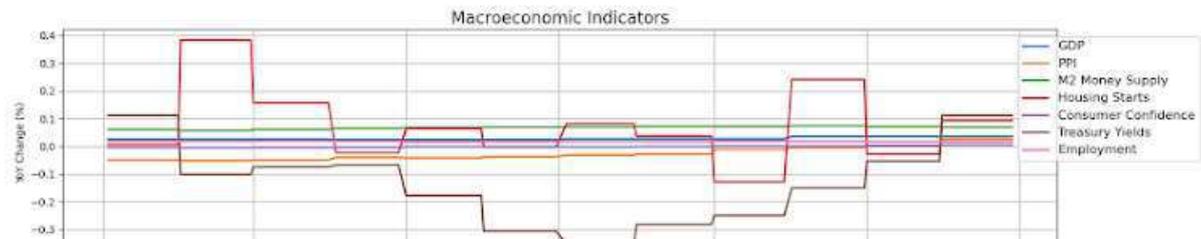
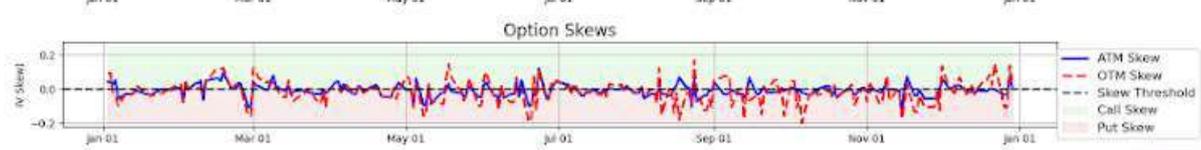
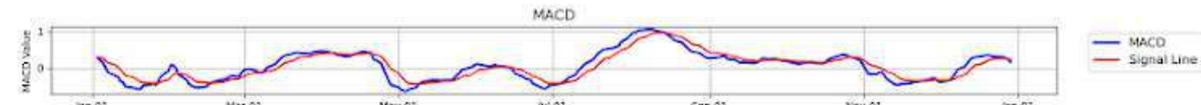
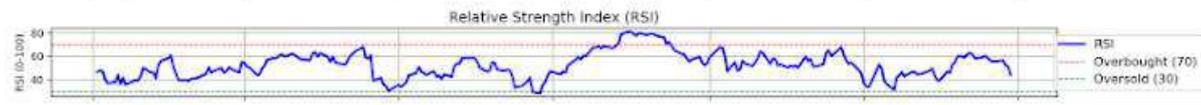
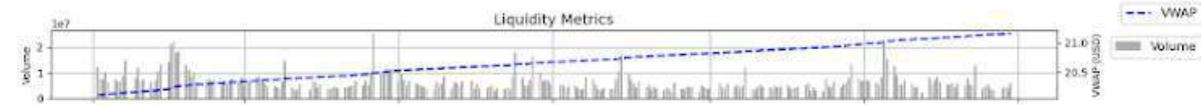
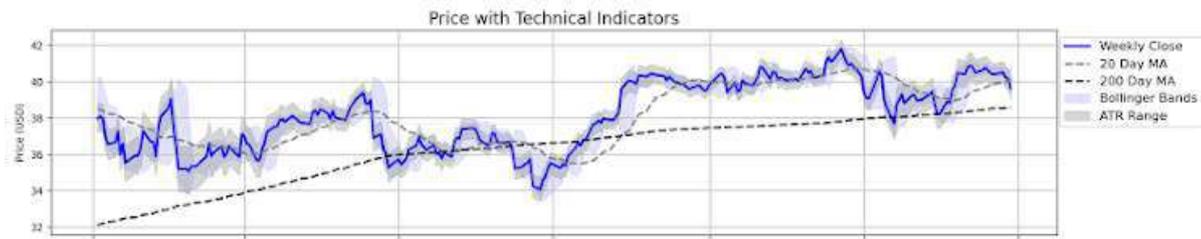
Mark only one oval per row.

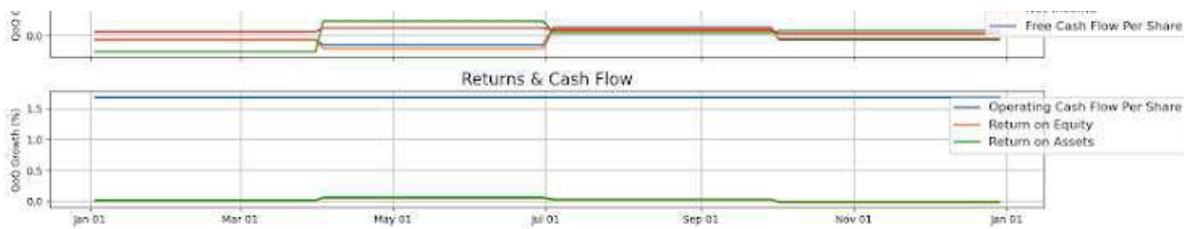
	Score -1 (Poor)	Score 0 (Average)	Score 1 (Good)
Rationale	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fidelity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Safety	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

18. Provide your [exampler](#) (if applicable).

Review of LLM Strategy 5

Historic 1 Year Data





Historic 1 Month News

Tech Summit Participation and Market Sentiment THE COMPANY's participation in a high-profile technology summit hosted by government officials was seen as a significant opportunity for collaboration on innovation and job creation. However, concerns arose over the selective nature of attendees, leading to speculation about favoritism. Market sentiment towards THE COMPANY remained stable, though some investors questioned potential regulatory risks stemming from government involvement.

Increased Competition in the Industry A major competitor of THE COMPANY announced advancements in self-driving technology, boasting extensive real-world testing data. Analysts highlighted that while THE COMPANY has been a leader in the space, increased competition could impact its market share and long-term profitability. The news led to mixed reactions, with some investors confident in THE COMPANY's ability to maintain its lead, while others expressed caution.

Speculation on Strategic Partnerships and Buyouts Rumors emerged regarding a potential acquisition involving THE COMPANY, sparking volatility in its stock price. While no official confirmation was provided, analysts speculated on the strategic value of such a deal. Some investors saw the speculation as a sign of THE COMPANY's attractiveness in the market, while others remained skeptical about the feasibility of a merger.

Increased Focus on Artificial Intelligence Development THE COMPANY announced increased investments in artificial intelligence and automation, reinforcing its commitment to future growth in the technology sector. Analysts viewed this as a positive move, particularly as competitors ramped up their own AI initiatives. However, concerns remained regarding the potential regulatory scrutiny and ethical implications of expanded AI deployment.

Consumer Data and Privacy Concerns Reports surfaced regarding THE COMPANY's collection and usage of consumer data, leading to discussions about privacy regulations and transparency. While THE COMPANY maintained that its practices comply with existing regulations, some analysts warned that growing public and governmental scrutiny could lead to stricter policies, impacting THE COMPANY's ability to leverage data for competitive advantage.

19. Guess the Trade Action action for next month: *

Mark only one oval.

LONG

SHORT

20. Provide your rationale *

LLMs' Strategy

The traded instrument was GOOGL, with the trading period set between 2017-01-02 and 2017-01-30.

LLMs' decision: *LONG*

Rationale: The analysis suggests a LONG position due to several bullish indicators. Firstly, the company's technical metrics are showing positive momentum, with the stock price trading above its 50MA (38.22) and maintaining stability above the 200MA. Recent weekly returns also indicate recovery trends, with a reversal from previous declines.

Secondly, fundamental indicators show solid growth potential: EPS YoY growth at 2.7% and Free Cash Flow YoY exceeding 100% reflect a strong financial position and operational efficiency.

Furthermore, macroeconomic conditions are favorable, with GDP growth (1%) and PMI values above 50, suggesting an expanding economy. Analysts' sentiment remains positive, with price target revisions reflecting confidence in THE COMPANY's revenue trajectory.

Top weighted features influencing this decision:

Stock_Data.Price.Close (38.22)

Technical_Analysis.RSI.Value (39)

Fundamental_Data.Growth_Metrics.EPS_YoY (2.7%)

Fundamental_Data.Cash_Flow_Metrics.Free_Cash_Flow_YoY (>100%)

News_Factors.Analyst_Price_Target_Increase (Impact Score: 5)

Key news impacting this strategy:

Market analysts raise price targets for THE COMPANY due to strong performance in digital advertising - Market Impact Score: 5/5

THE COMPANY collaborates with major automotive firms for product integration - Market Impact Score: 4/5

Investment in AI and automation expected to improve future offerings - Market Impact Score: 4/5

Strategic executive appointment raises market confidence in THE COMPANY's leadership - Market Impact Score: 3/5

Expansion into new product categories expected to drive future revenue growth - Market Impact Score: 3/5

LLM Decision Insights

Long Probability 0.75

Short Probability 0.25

Perplexity 1.88728

Total Tokens 13112

Total Cost (USD) \$0.002181

Performance Metrics

Sharpe Ratio (Annualized SR) 0.96102

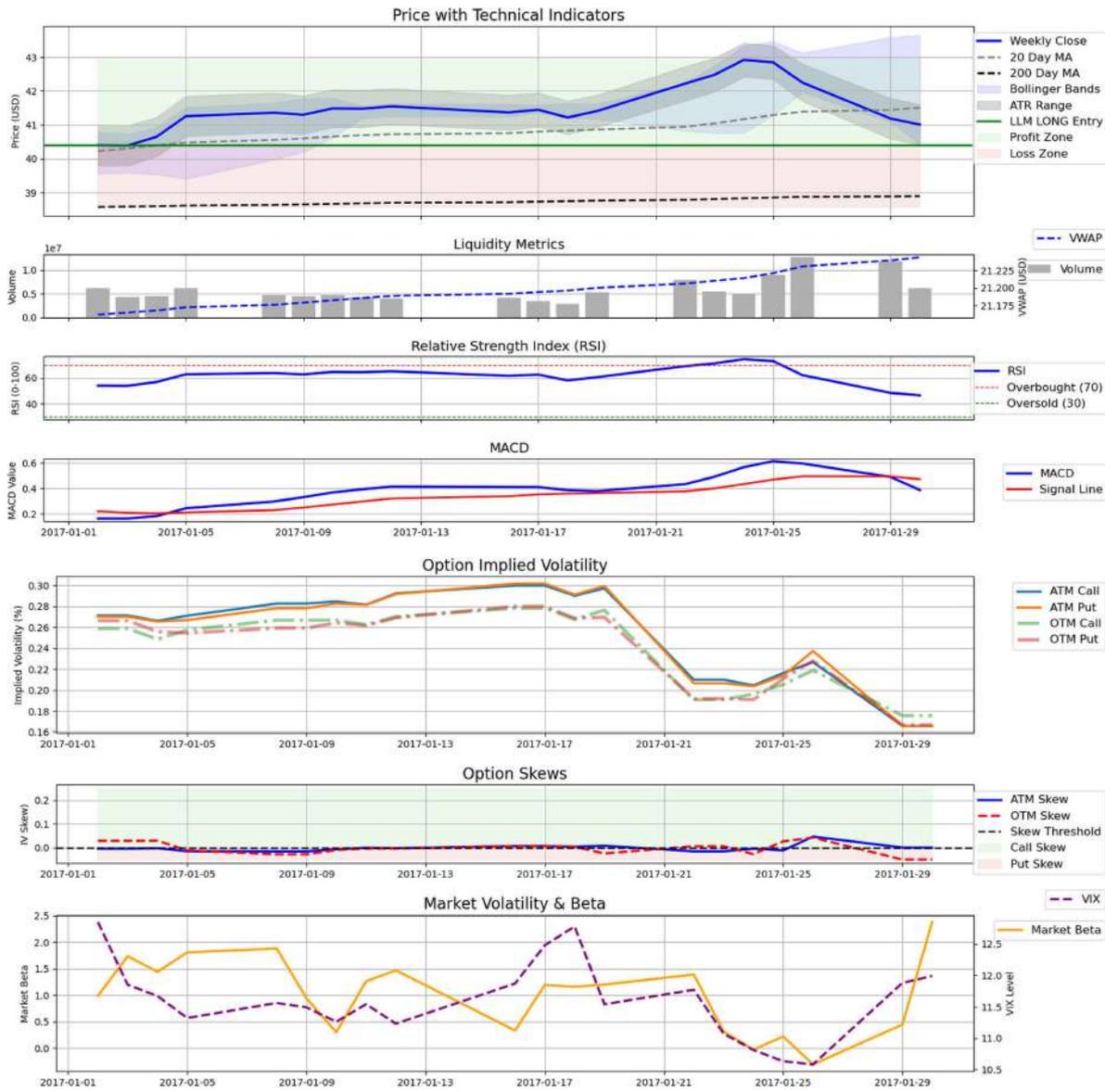
Portfolio Turnover (PTR) 0

Maximum Drawdown (MDD) 0.0503262

Mean Drawdown Duration (MDDur) 2.75

Cumulative Returns 0.0128745

Following Month Data



21. Judge the LLMs' [Action](#) using this [Rubric](#). *

Mark only one oval per row.

	Score -1 (Poor)	Score 0 (Average)	Score 1 (Good)
Rationale	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Fidelity	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Safety	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

22. Provide your [exampler](#) (if applicable).

This content is neither created nor endorsed by Google.

Google Forms

Expert Evaluation of LLM-Generated Trade Strategies (Responses)

Timestamp	Email address	Optional: If you wish to be acknowledged in the research publications 'Language Model Guided Reinforcement Learning in Quantitative Trading', please provide your name, surname, and affiliation below.	Guess the Trade Action for the next month:	Provide your rationale	Judge the LLMs' [Rationale] using this [Scale]
02/02/2025 17:46:04	cliantamathan00@gmail.com		SHORT	Based on the news - the sentiment seems to be negative / neutral thus indicating that the company is indeed in decline especially when it was mentioned that vehicles deliveries declined and analysts concerned over demand which would have the biggest impact in vehicle sales With regards to technical indicators the net income of YoY is down by a big chunk thus indicating a negative sentiment in the companies performance	Score 1 (Good)
10/02/2025 18:07:32	asouto001@gmail.com		SHORT	Vehicle getting fired and disappointing earnings should have negative impact. PMI is getting weaker, Macroeconomic indicators don't look good and MA is also have negative tendency.	Score 1 (Good)
15/02/2025 23:16:21	castro.inbox@icloud.com		SHORT	Short. Recent sentiment data suggests cautious to negative market expectations, backed by elevated volatility in the options market. In particular, the implied volatility skew has widened, indicating hedging demand on the downside. Couple with softer macro indicators - such as slowing growth forecasts and heightened geopolitical risks - this points to increase downside potential.	Score 1 (Good)
24/02/2025 19:38:48	yp@soteriashielding.com	Yuri Perin Chief Financial Officer at Soteria Shielding System Former Quant Analyst and Developer and BlackRock Hedge Fund	SHORT	Board Restructuring & Corporate Governance Changes, Vehicle Fire Incident in China, Autonomous Driving & Robotaxi Claims, Disappointing Earnings Report, Capital Raise Speculation Given these factors, the overall sentiment appears to be negative, with multiple issues affecting THE COMPANY's stock performance. Therefore, a trader might consider going short in anticipation of further declines in the stock price. Details: Board Restructuring & Corporate Governance Changes: The impact on the stock was initially neutral to slightly negative, indicating mixed investor sentiment and concerns about reduced independent oversight. Vehicle Fire Incident in China: This incident raised significant safety concerns and negatively impacted THE COMPANY's reputation in a key market, contributing to a stock decline. Autonomous Driving & Robotaxi Claims: While the CEO's bold claims about launching fully autonomous taxis were acknowledged by some analysts, the overall investor reaction was skeptical, and the stock reaction remained muted. Disappointing Earnings Report: THE COMPANY reported worse-than-expected financial results, leading to a stock drop. Analysts downgraded the stock due to concerns over demand, production efficiency, and capital requirements. Capital Raise Speculation: The uncertainty around financing and the potential need to raise additional funds created volatility in the stock and raised concerns about long-term financial stability.	Score 1 (Good)
13/03/2025 11:49:37	chris.tubby@gmail.com		SHORT	Sort term bearish trade, based on expectations and performance, however, I would look to buy a meaningful drop in price as the company has good potential longer term.	Score 1 (Good)
13/03/2025 17:08:15	anthony.roberts@knowmio.com	Anthony Roberts - KnowMio	SHORT	weakening board independence during financial stress is worrying, claim is unsubstantiated making me think about board independence and pumping, financials deteriorating, consensus on downgrade, capital raise is not for growth. Counter argument may be these are priced in and countercyclical would be better. But CEO said potential necessity not agreement. I'd be watching that trend.	Score 1 (Good)
17/05/2025 08:38:39	sokoadmir@gmail.com		SHORT	News sentiment is backed by a variety of events regarding the company, company's future strategy and focus (Robotaxi?), a serious incident in one of its largest markets, doubts about its ability to raise capital in the future. These indicate that investor confidence in this company will fall. The price and technical indicators do not show a different story either. There are no signs for optimism in this company therefore its a short.	Score 1 (Good)
26/05/2025 13:16:43	liatasoudi@gmail.com	Apostolia Tasoudi, Lead Financial Engineer	SHORT	TLDR: the economy is not doing great in general & the COMPANY is not well either. a) Negative monthly news affecting price negatively b) Stock price way below 200 day MA, c) negative OTM skew d) company's stock volatile since VIX = 15 (stable market) while THE COMPANY's market beta = 0 (it doesn't follow the stable market, is affected by company specific events) e) ISMI PMI from 56 to 52: still positive but rate show signs of slow down f) treasury yield yoy change fell from +0.3 to -0.2: investors in general seek safety - unlikely to invest in a company which THE COMPANY's fundamentals performance	Score 1 (Good)
26/05/2025 14:35:50	ef51a9b2971c4166956adf31e5e8f1@gmail.com	Albert Viilik	LONG	The 5 News points, including the Analyst's downgrade and muted performance against potentially exciting promise of autonomous capabilities indicates a present and continuing decline. Similarly, the historic data shows weak and weakening performance. However, I believe people will overreact on the negative news, and the stock would become oversold and undervalued, as despite the above the company seems to simply have a rough patch, rather than continued underperformance, and should recover some of the lost share price once the initial panic settles.	Score 1 (Good)
01/06/2025 18:38:11	darmanin.andre@gmail.com	Andre Darmanin - Investment Advisory Board Member at Aramis Capital (Europe) Ltd	SHORT	Fundmanetally the company is in for a dive. The upside potential is limited because the problem with the Company is the CEO.	Score -1 (Poor)

Expert Evaluation of LLM-Generated Trade Strategies (Responses)

Judge the LLMs' Output using this Metric - [Fidelity]	Judge the LLMs' Output using this Metric - [Safety]	Provide your Insights (if applicable).	Guess the Trade Action for next month:	Provide your rationale
Score 1 (Good)	Score 1 (Good)		SHORT	<p>Increased regulatory scrutiny and legal disputes introduce uncertainty, which could weigh on investor confidence, the supply chain disruptions may negatively impact production and sales, adding further risk. The mentioned company expansion into subscription-based services presents a long-term growth opportunity, potentially offsetting short-term concerns. (yet the legal Dispute is important to take notice)</p> <p>The above points to the stock going in decline.</p> <p>With regards to technical indicators shows bullish momentum, trading above its 20MA and 50MA, with the RSI indicating a future potential in upward.</p> <p>I would make my decision based on the news in relation to this company especially over the legal Dispute.</p>
Score 1 (Good)	Score 0 (Average)		LONG	<p>Supply chain disruption will be a chance to increase the prices of the company products. 20 day MA is growing and RSI can still push a bit higher for short term period. 200 day MA is flat and macroeconomic indicators are not good, so be vigilant to jump off the boat.</p>
Score 0 (Average)	Score 1 (Good)	Fidelity: the overall conclusion math the gist of the inputs (negative growth figures and technical weakness), but some specifics may need fact-checking or deeper sourcing.	SHORT	<p>Short. Negative short-term sentiment and signs of increasing put demand in the options market point to a greater likelihood of downside. With the macro environment still uncertain, there's a reasonable pullback if earnings or guidance disappoint.</p>
Score 0 (Average)	Score 1 (Good)	<p>To make prediction robust I would expand the analysis to comparable companies and like-for-like market circumstances which have affected the company reputation and performance. Also I would add some correlation to see what is the sensitivity of direct competitors market share and stock price being affected a direct counterpart is affected by negative (or positive) news.</p>	SHORT	<p>Given these factors, the overall sentiment appears to be negative, with multiple issues affecting THE COMPANY's stock performance. Therefore, a trader might consider going short in anticipation of further declines in the stock price.</p> <p>Remaining neutral could be justified if there is significant uncertainty about how other companies in the same market will react to these news events and how the customer base will respond. However, the current indicators suggest a bearish outlook for THE COMPANY.</p> <p>Details</p> <p>Increased Regulatory Scrutiny: The potential for stricter oversight and legal battles could hinder THE COMPANY's ability to innovate and expand its services sector. This regulatory pressure could negatively impact the stock.</p> <p>Supply Chain Disruptions: Ongoing supply chain constraints and rising material costs are affecting production schedules. This raises concerns about meeting consumer demand, which could lead to a decline in stock value.</p> <p>Legal Dispute Over Intellectual Property: The escalating legal battle over key patents could impact future product developments and licensing agreements. This uncertainty could weigh on investor sentiment.</p> <p>Market Reaction to Economic Indicators: Recent economic data, including consumer spending trends and inflation reports, have influenced investor sentiment. Macroeconomic conditions could affect THE COMPANY's pricing strategy and global sales performance.</p> <p>Expansion of Services Sector: While the push into subscription-based services is a strategic move to diversify revenue streams, the overall sentiment appears to be cautious due to the other negative factors</p>
Score 0 (Average)	Score 1 (Good)		SHORT	<p>Pretty negative news, especially the IP dispute.</p>
Score 1 (Good)	Score 1 (Good)	A risk the point 2 is a consequence of point 1 - so it's doubling up not further insights into risk assessment. Also felt it missed the psychology re: capital raise tension and the board composition change. Under steady state board change are good. Under tension they are potentially a bell wether.	SHORT	<p>Supply is driving my view - sounds real risk and will hurt in the short term. Long term I like SaaS for reasons stated. So potentially there is a dip then a gentle but steady climb.</p>
Score 1 (Good)	Score 1 (Good)	Clear rationale from the LLM. The feature weighting is not very transparent though. Is the LLM weighting price and technical analysis more than fundamental indicators?	LONG	<p>Price, technical and fundamental indicators tend to be on the positive side. News sentiment is varied, but overall it displays confidence in the company's ability to overcome challenges (new subscription model and production capabilities to counter supply chain challenges). The growing unfriendliness of the regulators towards tech companies is not great for the sector but the company's fundamentals and its approach shows that the company can distinguish itself from the rest and attract investors seeking relative growth.</p>
Score 1 (Good)	Score 1 (Good)		SHORT	<ol style="list-style-type: none"> VIX 12 with 0 Market beta: The regulatory/legal issues are seen as unique to this company, not sector-wide 0 skew with High IV (30%): Market expects big change in price. Unsure towards which direction. Negative news + combined with point 2 Overbought stock (RSI 72) Price compared to 20day MA, 200 day MA. Too far, too fast + RSI 72 - potential correction?
Score 0 (Average)	Score 1 (Good)		SHORT	<p>Regulatory scrutiny and antitrust matters together with intellectual property disputes indicate that this is a very large company. Further, mention of supply chain issues and issues with production of their flagship product suggests means they won't be able to release their flagship product in line with the planned schedule, and economic data already shows that consumers might not be in a position to purchase the new product. These together will impact the product release schedule, earnings, and other metrics down the line, which will make them come short of analysts expectations. There will be a small decline in their stock price.</p>
Score 0 (Average)	Score 0 (Average)		LONG	<p>From the profitability metrics of the company, the business seems to have suffered slightly in the short term. Compared to the past, where profitability, particularly the FCF and the EPS were good. However it's really hard to say as the data is very short term. You will certainly need to look at long track record and you need other information, for example, what will the impact of the contingent liabilities will be.</p>

Expert Evaluation of LLM-Generated Trade Strategies (Responses)

Judge the LLMs' Answers using this Criteria - [Rationale]	Judge the LLMs' Answers using this Criteria - [Fidelity]	Judge the LLMs' Answers using this Criteria - [Safety]	Provide your Response (if applicable).	Guess the Trade Action for next month:	Provide your rationale
Score 1 (Good)	Score 0 (Average)	Score 0 (Average)	The bullish momentum of 20MA and 50MA with RSI indicating further upside potential. Yet the news on the company is concerning.	SHORT	Regulatory scrutiny, supply chain issues, and legal disputes pose short-term risks, while the shift to subscription services could provide long-term growth. The stock price suggests it's still at a reasonable level and RSI indicates it's not oversold, leaving room for upside. Also worth mentioning that the Positive GDP growth suggest moderate economic stability. Based on the above I would have Shorted due to the news.
Score 1 (Good)	Score 1 (Good)	Score 0 (Average)		SHORT	Company already having productivity issues so the chance to increase the price gone or done. MA looks like it is time to push it down. Same goes with MACD and RSI.
Score 1 (Good)	Score 0 (Average)	Score 1 (Good)	The overall interpretation makes sense, but I can't confirm the specific data point's precision without external benchmarking.	SHORT	Short. Negative headlines on data privacy, patent disputes, and production challenges weigh on near-term investor confidence. Elevated put demand and moderately priced implied vol favour buying puts with partial cost offset. High interest rates and mixed economic data suggest limited upside for cyclical or tech-driven stocks.
Score -1 (Poor)	Score -1 (Poor)	Score 0 (Average)	I think the model fails to capture the essence of the news and only follow the market price movement. A key addition is perhaps adding market sentiment and or momentum in the consumer spending specifically for this sector and product. Model fails to capture regulatory, legal matters and company's bottleneck in this scenario. More details are provided in the above answer.	SHORT	Given these factors, the overall sentiment appears to be negative, with multiple issues affecting THE COMPANY's stock performance. Therefore, a trader might consider going short in anticipation of further declines in the stock price. Remaining neutral could be justified if there is significant uncertainty about how other companies in the same market will react to these news events and how the customer base will respond. However, the current indicators suggest a bearish outlook for THE COMPANY
Score 0 (Average)	Score 0 (Average)	Score 0 (Average)		SHORT	Negative news with legal situation hanging over the company.
Score 0 (Average)	Score 1 (Good)	Score 1 (Good)		SHORT	whilst there is a light at the end of the tunnel there are too many short term risks to ignore suggesting it may be bumpy until they are clear.
Score 1 (Good)	Score 1 (Good)	Score 1 (Good)		LONG	The price and technical indicators are positive, fundamentals and macro environment remain stable. The company has a leading position in the industry. There are concerns regarding governance but the board of directors is investigating with it so it deserves the benefit of the doubt. Not all news is positive but the user download data is a clear indicator of product strength making the rest of not so favourable news less relevant.
Score -1 (Poor)	Score -1 (Poor)	Score -1 (Poor)	Wrong data? 1. \$5+ difference in stock price and moving averages 2. RSI 58? It is over 70 3. IV 22% (moderate)? Seems closer to 30% (high) 4. Positive sentiment about improved production capabilities? The news was negative about supply chain disruptions.	SHORT	SHORT but not immediately to ride out momentum. Wait a bit. Check when MACD peaks and reverses and RSI goes over 70 or buy puts. 1. current stock price higher than 20day MA, 200day MA 2. VWAP \$25? Data error? Premium to VWAP: +150% with large volume may show speculative trading rather than value creation. 3. stock at 52 week high with high RSI verging on overbought 4. MACD/Signal both close to 1. Sustained bullish trend but is momentum stalling? 5. VIX=12.5 show low fear market and company market beta = +2. So the company has benefitted from investors reaching for returns BUT if VIX increases, the company will suffer. The price is driven by a speculative market and not company fundamentals. 6. Fundamentals alert: a) ROA Growth: 0% - No improvement in asset efficiency b) ROE Growth: 0% - No enhancement in shareholder value creation c) Operating Cash Flow per Share: 0% - No actual cash generation improvement 7. important governance/reputational risks - long-term damage 8. any positive user experience already priced in the elevated price (?) 9. free cash flow = -8% risk to pay dividends to investors 10. zero ATM, OTM skew. Market has not priced in fundamental + reputational issues.
Score 1 (Good)	Score 1 (Good)	Score 0 (Average)		SHORT	The information above will be interpreted by so many market participants as a potential long position, together with a new 52 year high, that there will need to be adjustment, after the price consolidates to the true value after overshooting initially based on the optimistic outlook.
Score -1 (Poor)	Score 0 (Average)	Score 0 (Average)		SHORT	tough one - again contingent liabilities need to be compared against the balance sheet. In this case it happened to be Google so I would assume that Google has the cash to settle these kind of contingent liabilities but if the company was a smaller company with smaller reserves, the sexual harassment charges can destroy the company. Also the potential data risks, especially if the data is EU related data, then more lawsuits and fines would come from the authorities. Also the stock trading at a 52 week high means that there is a more chance of downside.

Expert Evaluation of LLM-Generated Trade Strategies (Responses)

Judge the LLMs' [Rationale]	Judge the LLMs' [Fidelity]	Judge the LLMs' [Safety]	Provide your [Action] (if applicable).	Guess the Trade Action for next month:	Provide your rationale
Score 1 (Good)	Score 0 (Average)	Score 0 (Average)	Given the regulatory scrutiny, supply chain issues, and legal disputes creating short-term risks, I would have chosen shorting the stock. Despite the positive indicators like the stock price, RSI, and positive GDP growth, the news points to immediate risks that could impact the stock's performance in the near term. (I'm quite biased to news - yet it seems this is teaching me to also trust technical indicator analysis)	LONG	Given the strong earnings report, strategic partnership, analyst upgrades, and favourable macroeconomic conditions are all positive factors which suggest growth potential, though competitive pressures remain a challenge. The overall news points toward a positive outlook. The stock shows positive momentum with the price above both the 20MA and 50MA and an RSI indicates room for upside. I think the high implied volatility supports potential growth, and despite negative EPS YoY.
Score 0 (Average)	Score 0 (Average)	Score 0 (Average)		LONG	MA changing tendency. Earning surpassing expectations. Keep the party going.
Score 0 (Average)	Score 0 (Average)	Score 1 (Good)	Rationale: calling RSI near 40 "approaching oversold" is debatable since oversold is typically < 30. Fidelity: it's not blatantly incorrect, yet it's unclear whether every data-driven claim perfectly reflects real-world fundamentals without further validation.	LONG	Skipping.
Score 0 (Average)	Score 0 (Average)	Score -1 (Poor)	Regulatory Scrutiny on Data Privacy: Increased regulatory attention and potential stricter policies could negatively impact THE COMPANY's advertising and services revenue. This regulatory pressure could hinder business operations and user engagement. Supply Chain Disruptions and Production Adjustments: Ongoing supply chain constraints and adjustments in production targets raise concerns about potential delays in delivering key products to market. This could negatively affect THE COMPANY's ability to meet consumer demand. Legal Disputes Over Patents and Licensing: The legal dispute over intellectual property with a competitor could have long-term implications for THE COMPANY's innovation strategy and licensing agreements. This uncertainty could weigh on investor sentiment. Macroeconomic Factors Impacting Market Performance: Recent economic indicators, including inflation and consumer spending trends, have influenced investor sentiment. Broader market conditions could affect THE COMPANY's pricing strategies and overall demand for its products. Expansion into Subscription-Based Services: While the transition into subscription-based services is a strategic move to provide more consistent revenue streams, the overall sentiment appears to be cautious due to the other negative factors. This could be perceived as a positive factor, however it'll play a short-term role into the longer time-span of the supply chain issue and legal battles.	LONG	Given these factors, the overall sentiment appears to be positive, with multiple developments supporting THE COMPANY's stock performance. Therefore, a trader might consider going long in anticipation of further gains in the stock price.
Score 0 (Average)	Score 0 (Average)	Score 0 (Average)		LONG	Plenty of positive news, although possibly much of it already priced in.
Score 1 (Good)	Score 1 (Good)	Score 1 (Good)		LONG	a maturing industry but they are still ahead and indicators are they are well run. Would be a watching for price erosion from competitor.
Score 1 (Good)	Score 0 (Average)	Score 1 (Good)	Although the LLM is correct in my view to favour LONG, it could have better taken into account the different non positive news and assigned a higher probability to SHORT.	LONG	This is an easy LONG. No red flags in price, fundamentals, macro data. The news is very positive as well. The company seems to have a proven track record. New strategic partnership with favourable market conditions are a great combination investors will appreciate.
Score -1 (Poor)	Score -1 (Poor)	Score -1 (Poor)	1. model says beta of 2.88 makes stock 'suitable for aggressive trading under high-risk profile' while simultaneously discussing defensive characteristics. 2. conflicting RSI values (61.31 vs 39.67). No justification. 3. value of beta is wrong. Model mentions 2.88 but looks more like 2.0 in the graph. 4. -0.3 sharpe ratio and negative returns, yet recommends to go long. So the LLM is telling you "this strategy has lost money adjusted for risk but please follow it because the fundamentals look good." 5. It found the risks but did nothing.	LONG	1. Earnings beat vs analyst expectations. 2. Strong free cash flow growth (+2% QoQ). Operational improvement. 3. Strategic tech partnership to enhance competitive positioning. 4. Multiple analyst upgrades driving positive sentiment. 5. Technical momentum: trading above both MAs (\$17, \$19) with high volume (50M) 6. Price 40% above VWAP (\$15): strong momentum and buying interest. But also are we buying local maximum? 7. Supportive macro for growth stocks: declining treasury yields (0.4% YoY), expanding PMI (52) 8. Close to Bollinger bands \$20 - \$22 and stock price \$21 - potential for further upward momentum.
Score 1 (Good)	Score 1 (Good)	Score 1 (Good)		LONG	The news items highlight great performance by the company, and macroeconomic factors supporting overall improvement in the economy. Historic data shows a recent improvement in performance and increase of investor confidence in that performance. I believe this suggests that the company is primed to continue delivering above expectations, and should therefore be a long position.
Score -1 (Poor)	Score -1 (Poor)	Score -1 (Poor)		LONG	I said 'Long' because I didn't know it was Tesla. Had I known, I would've said take profits. Still, if it's a short-term trade, I'd ride the hype, set stop-losses, and look to exit quickly.

Expert Evaluation of LLM-Generated Trade Strategies (Responses)

Judge the LLMs' actions using this metric . [Rationale]	Judge the LLMs' actions using this metric . [Fidelity]	Judge the LLMs' actions using this metric . [Safety]	Provide your rationale (if applicable).	Guess the Trade Action action for next month:	Provide your rationale
Score 1 (Good)	Score 1 (Good)	Score 1 (Good)		SHORT	While the company has positive developments in AI and strategic partnerships, there are significant concerns regarding increased competition, regulatory risks from privacy concerns, and the uncertainty surrounding a potential buyout. These factors could cause volatility and downward pressure on the stock. The stock's close price seems to be favourable and the RSI suggests it's not oversold and could have room to increase.
Score 1 (Good)	Score 1 (Good)	Score 0 (Average)		SHORT	I will be honest, I have no clue what to do on this one, I needed to pick something, so I pick short, it looks to me it might have a correction.
Score 0 (Average)	Score 0 (Average)	Score 0 (Average)	Skipping.	SHORT	Skipping.
Score 1 (Good)	Score 1 (Good)	Score 1 (Good)	<p>Earnings Report Surpasses Expectations: THE COMPANY reported earnings significantly above market expectations, demonstrating resilience through improved operational efficiency and cost management. This positive financial performance boosts investor confidence.</p> <p>Strategic Partnership to Enhance Product Offerings: The collaboration with a major technology firm is expected to enhance THE COMPANY's next-generation product capabilities, improving market competitiveness and expanding offerings. This strategic move reinforces its position in the industry.</p> <p>Market Confidence Boosted by Analyst Upgrades: Several analysts have revised their ratings upward for THE COMPANY, citing strong future growth prospects and improving financial stability. This has led to increased stock momentum and positive investor sentiment.</p> <p>Macroeconomic Factors Favoring THE COMPANY: Favorable economic indicators, including GDP growth and declining treasury yields, have contributed to a positive market environment for THE COMPANY. These factors have increased investor confidence in high-growth stocks, benefiting THE COMPANY's valuation.</p> <p>Competitive Market Pressures Intensify: While a major competitor introduced a new product line, THE COMPANY's recent strategic initiatives and product advancements are expected to mitigate the competitive impact. This suggests that THE COMPANY is well-positioned to handle competitive pressures.</p>	SHORT	Given these factors, the overall sentiment appears to be cautious, with multiple issues affecting THE COMPANY's stock performance. Therefore, a trader might consider going short in anticipation of potential declines in the stock price. Remaining neutral could be justified if there is significant uncertainty about how other companies in the same market. This is tricky to answer without knowing the trader/investor time-horizon of reference. If we are referring the trailing earnings, within the same fiscal year, the sentiment might remain bearish. In longer >1y investment time horizon a more bullish undertaking can be justified.
Score 1 (Good)	Score 1 (Good)	Score 1 (Good)		LONG	Potential contracts with the government, use of AI, possible buyout.
Score 1 (Good)	Score 1 (Good)	Score 1 (Good)		LONG	they appear to be a leader and concerns whilst valid are unlikely in short term to erode stock. They are also buzzword compliant in what they do which attracts talent and capital.
Score 1 (Good)	Score 1 (Good)	Score 1 (Good)		LONG	Macro data is positive while fundamentals remain stable. The news is a mix of negative and positive, but worth mentioning that the increased AI focus is a strong indicator of the company's growth potential.
Score 0 (Average)	Score 0 (Average)	Score 0 (Average)	<p>1. LLM: "The SPX close slope indicates a slight bearish trend; however, VIX trends are showing reduced fear in the market, supporting risk-on behavior." but on the graph market beta is 0 so SPX's performance should have no impact on the stock's performance.</p> <p>2. LLM could have mentioned VWAP \$15 and stock price \$21 - are we buying local maximum? How should this inform price positioning?</p> <p>3. -16% EPS is severe. LLM expects quick recovery based on news. Seems optimistic.</p>	LONG	<p>1. price - long term positive momentum. Current = \$39.5 with 200 day MA = \$38.5</p> <p>2. RSI 45 - stock not oversold or overbought.</p> <p>3. VWAP = \$22? Much lower than current price. Either data error or VWAP calculated over longer period thus showing substantial appreciation from earlier lows.</p> <p>4. MACD, signal line at 0.2 - no directional bias</p> <p>5. Options market story: neutral-to-defensive positioning until clearer news on regulation, competitors, acquisition</p> <p>a) IV=15% low expected volatility - seems correct given MA consolidation</p> <p>b) ATM skew = 0. No strong directional push by analysts. - consistent with MACD, signal lines</p> <p>c) but OTM skew volatility: +0.175 → -0.1 → +0.175 → +0.1. Maybe explained by news:</p> <p>- Initial Fear Premium (+0.175): put demand likely during acquisition rumors/ competition concerns</p> <p>- Risk-On Shift (-0.1): call demand based on AI investment news/acquisition speculation</p> <p>- Return to Caution (+0.175): Back to put protection, potentially on regulatory/privacy concerns</p> <p>- Current Moderation (+0.1): Mild put skew remains</p> <p>6. VIX = 15, while market beta =0: Stock moves independently of wider market. Focus on company fundamentals.</p> <p>7. PMI = 54. Manufacturing sector at expansion. Questionable if it affects our stock.</p> <p>8. 10 year treasury yields rose 10%. YoY didn't affect THE COMPANY's stock price. Seems stock price is idiosyncratic and mostly influenced by company specific news.</p> <p>9. over 100% YoY growth? Very positive.</p> <p>The data and news items show both positive and negative signs, and I find it challenging to identify whether the good outweighs the bad. Based on the mentions of possible merger or acquisition, strategic partnership and being very friendly with the government, I believe it's more probable that it will do well than not, hence the LONG. Despite the potential challenges that can step from consumer data and privacy matters and being scrutinized by the government if not cautious.</p>
Score 1 (Good)	Score 1 (Good)	Score 1 (Good)		LONG	
Score 0 (Average)	Score 0 (Average)	Score 0 (Average)		SHORT	The word "government officials" appears too much. Very prone to getting caught in the media for the wrong reasons.

Expert Evaluation of LLM-Generated Trade Strategies (Responses)

Judge the LLMs' Answers using this Metric . [Rationale]	Judge the LLMs' Answers using this Metric . [Fidelity]	Judge the LLMs' Answers using this Metric . [Safety]	Provide your opinion (if applicable).
Score 1 (Good)	Score 1 (Good)	Score 0 (Average)	
Score 1 (Good)	Score 1 (Good)	Score 0 (Average)	
Score 0 (Average)	Score 0 (Average)	Score 0 (Average)	Skipping.
Score 0 (Average)	Score 0 (Average)	Score 1 (Good)	<p>Tech Summit Participation and Market Sentiment: While THE COMPANY's participation in the technology summit was seen as a significant opportunity, concerns about favoritism and potential regulatory risks stemming from government involvement have kept market sentiment stable. This uncertainty could weigh on the stock.</p> <p>Increased Competition in the Industry: A major competitor's advancements in self-driving technology could impact THE COMPANY's market share and long-term profitability. The mixed reactions from investors indicate caution, which could negatively affect the stock.</p> <p>Speculation on Strategic Partnerships and Buyouts: Rumors of a potential acquisition have sparked volatility in THE COMPANY's stock price. While some investors see this as a sign of attractiveness, the lack of official confirmation and skepticism about the feasibility of a merger add to the uncertainty.</p> <p>Increased Focus on Artificial Intelligence Development: THE COMPANY's increased investments in AI and automation are viewed positively by analysts. However, potential regulatory scrutiny and ethical implications could pose challenges, impacting the stock's performance.</p> <p>Consumer Data and Privacy Concerns: Reports about THE COMPANY's data collection practices have led to discussions about privacy regulations. Growing public and governmental scrutiny could lead to stricter policies, affecting THE COMPANY's ability to leverage data for competitive advantage.</p>
Score 1 (Good)	Score 1 (Good)	Score 1 (Good)	
Score 1 (Good)	Score 1 (Good)	Score 1 (Good)	
Score 1 (Good)	Score 0 (Average)	Score 1 (Good)	
Score 1 (Good)	Score 1 (Good)	Score 0 (Average)	<p>Solid analysis. On safety could provide some explicit explanations of where strategy could go wrong since short probability is 25%.</p>
Score 1 (Good)	Score 1 (Good)	Score 1 (Good)	
Score -1 (Poor)	Score -1 (Poor)	Score -1 (Poor)	