COMP0057 Literature Review: Training a Large Language Model with GRPO to create profitable cryptocurrency transactions given ’on-chain’ and ’CEX’ data.

Mohammed Khalil[1]

UCL, mohammed.khalil.20@ucl.ac.uk

Abstract. The following review examines critically the intersection of Large Language Models (LLMs), Reinforcement Learning and Cryptocurrency Markets. Further, it examines opportunities to leverage GRPO to train LLM based trading agents. To begin, structural characteristics of on-chain and CEx markets are surveyed with unique market features such as decentralisation, transparency and 24/7 availability highlighted. Following this, RL methodologies applied to financial markets are reviewed are reviewed and GRPO is introduced. The capabilities of LLMs in financial environments are explored, particularly the different "trader" and "alpha miner" roles. Finally, representative implementations are examined, such as FinMem’s layered memory trading agent, AlphaGPTs human alpha mining framework and a RL portfolio management framework. The review lays the groundwork for integrating cryptocurrency data into GRPO trained LLMs, outlining avenues for future development of profitable trading strategies.

1 Introduction

Cryptocurrencies, such as Bitcoin, Ethereum and Solana have gained significant traction over the past decade. They offer an alternative to the traditional financial system with decentralised markets, 24/7 trading and high volatility. The level of transparency distinguishes DeFi markets, as transactions on public blockchains are recorded on a shared ledger and available to all market participants [3]. This is very different to the traditional equity markets where trade data is dispersed among exchanges and often opaque. On-chain metrics, which have been described as an "untapped source of potential alpha" can be used to gather real-time insights into money flows, participant behaviour and even investor sentiment [14].

Centralised exchanges still however play a pivotal role in cryptocurrency trading, with the bulk of trading volumes still flowing through CEXs such as Coinbase and Binance. CEXs typically provide deeper liquidity and lower transaction costs relative to onchain transacting, attracting both retail and institutional usage. Similar to exchanges in equity markets, CEX’s maintain off-chain orderbook systems. As a result, both on-chain and off-chain exchanges must be monitored to get a fill picture of market supply and demand dynamics.

When crypto is mentioned in the news lately, it is usually due to the so called memecoin mania [1], with even the US President Donald Trump launching his own ’TRUMP’ token [4]. Valuations of such memes can be swayed dramatically by trends on Twitter, Tiktok, reddit and other platforms.

The confluence of on-chain transparency, 24/7 trading, and instant online sentiment means traders face a firehose of realtime data which human analysis alone struggles to digest. As a result, algorithmic and automated strategies have significant potential in crypto markets.

LLMs have emerged as powerful tools in several spaces as of late, including the financial sector. They are capable of parsing vast amounts of unstructured text to generate coherent responses, opening up novel applications in finance[31]. Unlike traditional quant methods that rely purely on numerical data, LLMs can incorporate both numerical and written data (such as news articles, balance sheets in equities, earnings transcripts, social media posts etc). The ability to understand context has significant applications for finance, especially in a narrative driven market such as crypto.

In parallel with advancements in LLMs, Reinforcement Learning (RL) models have been established as key in financial decision making. RL involves autonomous agents learning to make a sequence of decisions by firstly interacting with an environment then receiving feedback in the form of rewards[16]. In the context of trading, this could be executing transactions (buy/sell/hold) and being rewarded with profit. RL has been applied in several prior research articles in a trading context with success [11].

Finally, Group Relative Policy Optimisation (GRPO) is a recent innovation in the RL domain created by the team behind Deepseek, which aims to efficiently train large models (like LLMs) on complex tasks. [22]

The literature review critically examines the research encompassing the current state of Crypto, LLMs and RL to develop strong approaches towards solving the problem of training a model to create profitable transactions trading crypto.

2 Cryptocurrencies

Cryptocurrencies trade on a decentralised, 24/7 global market [8]. The continuous and borderless trading environment results in extreme volatility and structural breaks significantly exceeding what is found in traditional equity markets [2]. Prices for both majors (Top 10 tokens) and the remaining millions of tokens can and do swing dramatically, influenced by factors including but not limited to investor sentiment, network activity, macro news, 'memetic' narratives and marketing trends. Despite growing participation from retail investors and major institutional names such as Jump [18], Jane street [17], Citadel [5] and more.

In recent times memecoins, which can be defined as viral internet memes with little intrinsic utility, have surged in popularity due to speculative hype and coordinated pump and dump cycles [19]. This was further accelerated on the Solana blockchain through platforms such as 'pump.fun' which has enabled 'fair-launch' creation of hundreds of thousands of coins [13]. The use of 'on-chain' analytics indicates that monitoring large wallet activity is often an effective signal of impending momentum, hence many traders use tools to monitor 'whale' wallets [27]. Social media metrics are also incredibly pivotal. X (formerly Twitter) engagement is a significantly important (and also well researched) factor. [6] demonstrate that during the pandemic, Twitter engagement had a huge effect on 'Dogecoin' returns but little effect on Ethereum. This indicates that memecoin valuations are largely driven by the effects of social media virality.

[21] suggest that combining real-time onchain data with social sentiment trends can lead to forecasting of "runners" (memecoins with potential for rapid growth) and guide entry/exit levels.

3 Reinforcement Learning in Financial Environments

Reinforcement learning (RL) provides a sequential decision making framework, naturally suitable for trading agents that must continuously decide whether to buy, sell or hold a token. [10]. In a financial context, we see RL agents interact with environments such as market simulations to maximise cumulative rewards, typically profit or risk-adjusted profits. A very popular approach to RL in finance is policy gradient methods [10], which include 'actor-critic' algorithms and variants like Deep Deterministic Policy Gradient (DDPG) for continuous action spaces. Often used is Proximal Policy Optimisation (PPO), which is a policy on policy gradient algorithm used by OpenAI in 2017 [7]. PPO restricts policy updates within a trust region using a clipped surrogate objective, ensuring stable training and preventing a collapse in performance. Due to relative simplicity and reliability, it has become the baseline in many domains, including finance. For RL methods to work in noisy environments (such as finance), they require huge sample sizes and careful tuning. Further, they tend to overfit historical data without rigorous validation [15].

Diving deeper into Reinforcement Learning Applied to Trading Systems: A survey [10], we see Felizardo et al undertake an examination of how RL has been applied to trading tasks. The authors contribute:

– A workflow pipeline that reflects core RL frameworks to catagorise existing works.
– In-depth comparison of 29 articles along the developed pipeline
– Identification of trends in two categories - Single Asset Trading & Portfolio Optimisation.
– Insights and recommendations to inform future research

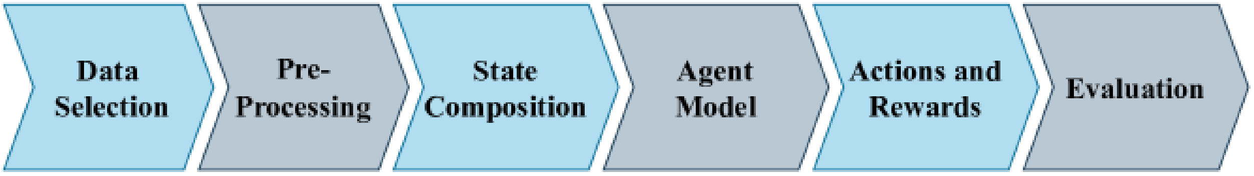


Fig. 1. Six key design dimensions for classifying RL trading systems

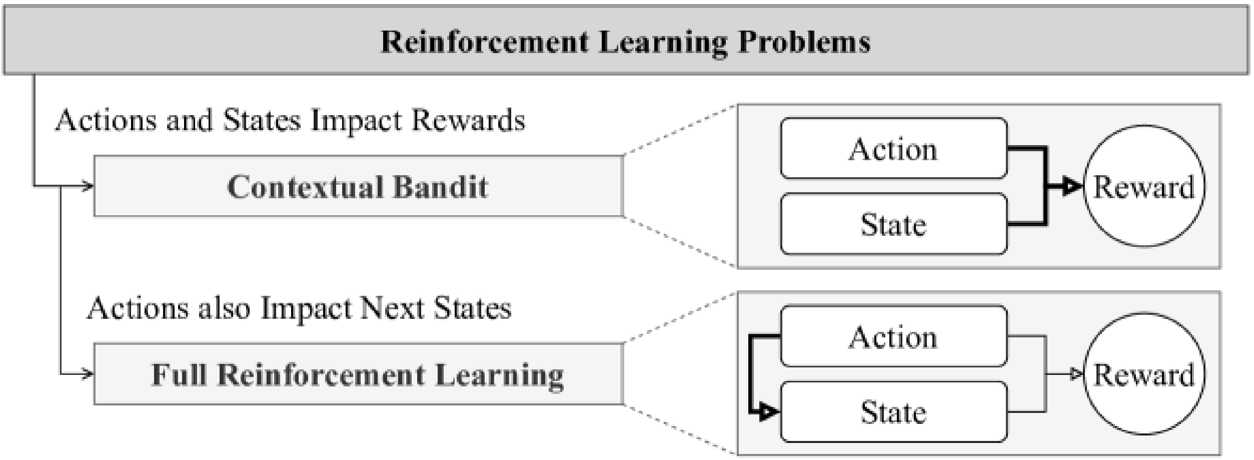

Reinforcement Learning Problems
Actions and States Impact Rewards
Contextual Bandit
Action
State
Reward
Actions also Impact Next States
Full Reinforcement Learning
Action
State
Reward

Fig. 2. Contextual Bandit vs Full Reinforcement Learning approaches in trading systems

The authors reviewed 29 works with each paper classified according to six key design dimensions. This is shown in the figure below:

RL problems are split into Contextual Bandit and Full Reinforcement Learning problem types

R S Sutton et al in Reinforcement Learning: An introduction [26] clearly highlight the difference. For the contextual bandit setting, also known as associative search, the agent has to undergo a sequence of independent "bandit" tasks. On each round it observes a context and selects an action which automatically results in a reward. There is NO state to state transitions. The overall goal is to learn a policy that maps each context to the action resulting in the highest expected reward yielded. For full RL, the problem is a closed loop interaction between the agent and environment. At each timestep, the agent senses the current state, acts and then receives a potentially delayed reward. The most important thing to note is that actions not only affect immediate rewards but also future states and subsequent rewards. The full RL agent has the goal of learning over time via trial and error without direct instruction a policy to maximise cumulative reward.

Technical indicators such as moving averages and RSI dominate in popularity with few models integrating fundamental features like sentiment embeddings.

Comparing Single asset and Portfolio studies, single asset studies use discrete actions -> Long, Short, Hold whereas Portfolio models employ continuous weight vectors characterised by $w = (w_1, w_2, \ldots, w_n)$, where w_i is the weight of the i-th asset of the portfolio comprising n assets.

Model rewards are defined as both 'real' profits and risk adjusted metrics such as Sharpe or Calamar ratio. Such ratios provide a way to define profit per unit of risk. Sharpe for example is,

$$\text{Sharpe Ratio} = \frac{E[R_p] - R_f}{\sigma_p}, \tag{1}$$

$E[R_p]$ represents portfolio expected return, R_f risk free rate, and σ_p standard dev . It was originally introduced as the "reward-to-variability" measure by [24] and then formalised as the *Sharpe Ratio* in [25].

4 GRPO

Group Relative Policy Optimisation (GRPO) is a recently proposed PPO variant aimed at more efficient training of large models, specifically LLM reasoning through the DeepSeekMath Paper by Shao et al [22].

To begin, we explore Proximal Policy Optimisation. Introduced by Shchulman et al. [23], It is an on-policy gradient method that balances a reliable way to improve policy with sample efficiency and efficient implementation. Unlike vanilla policy gradient that performs a single gradient step per batch of collected data, PPO maximises a 'clipped surrogate objective', which prevents policy updates from being overly large.

The below figure shows the original TRPO "surrogate" objective

$$r_t(\theta) \;=\; \frac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{\text{old}}}(a_t \mid s_t)}, \quad r_t(\theta_{\text{old}}) = 1$$

$$L^{\text{CPI}}(\theta) \;=\; \hat{\mathbb{E}}_t\big[r_t(\theta)\,\hat{A}_t\big] \;=\; \hat{\mathbb{E}}_t\left[\frac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{\text{old}}}(a_t \mid s_t)}\,\hat{A}_t\right]$$

Fig. 3. TRPO "surrogate" objective (CPI): the probability ratio $r_t(\theta)$ and its use in the clipped surrogate loss.

CPI refers to 'conservative policy iteration', however PPO modifies this objective to penalise changes to the policy that moves the probability ratio away from 1. This is shown in the below figure, where where ϵ is a small positive hyperparameter. The clipping operation ensures that updates that push the ratio above 1 are penalised.

$$L^{\text{CLIP}}(\theta) = \hat{\mathbb{E}}_t\Big[\min\Big(r_t(\theta)\,\hat{A}_t,\ \text{clip}\big(r_t(\theta),\, 1-\epsilon,\, 1+\epsilon\big)\,\hat{A}_t\Big)\Big] \tag{2}$$

Fig. 4. The clipped surrogate objective used in PPO, which limits policy updates to within ϵ of the old policy.

The change between PPO and GRPO are demonstrated in the Deepseek math paper in the following figure:

GRPO removes the need for additional value function approximation in PPO. The average reward of multiple sampled outputs is used instead, which are produced in the same question as the baseline.

The important result that this is a very efficient RL model, especially in the usecase of optimising LLMs for tasks with complex high variance rewards. It is for this reason that it has potential to be effective in financial market context.

5 LLM's in Financial Environments

Large Language Models are advanced AI models, based on the google transformer paper [28], trained on incredibly large text sources to generate and understand natural language. Modern LLMs like GPT4 have billions of parameters and learn patterns in human language in order to produce text which is coherent and even perform tasks of reasoning. Architecturally, LLMs rely on self attention mechanisms that allow them to process text in out in parallel and to capture long range dependencies[12].

Through learning from a huge variety of datasets crawled from the internet including books, articles, code etc, LLMs build a statistical model of language that can be adapted in many ways.

They can understand unstructured data (freeform text), but also handle unstructured inputs by treating them as text (for instance serialising tables or JSON).

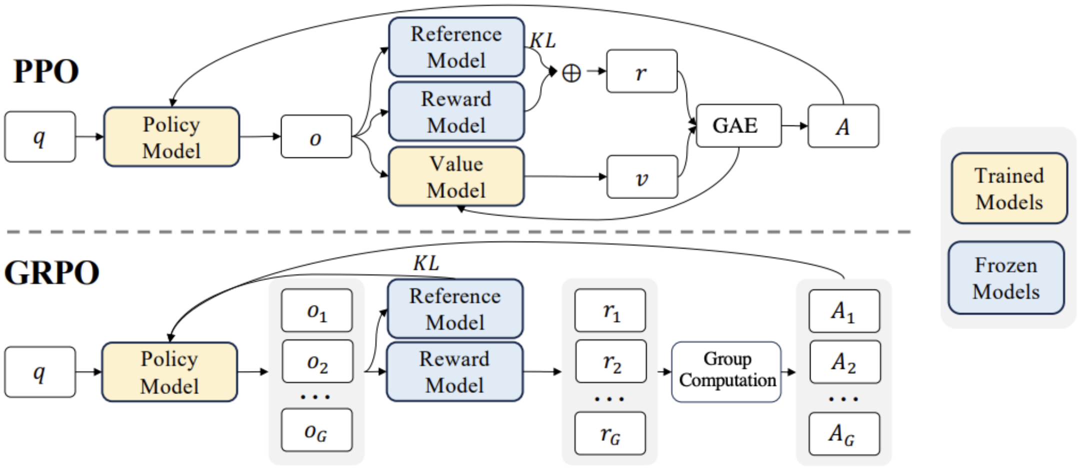


Fig. 5. Comparison between PPO and GRPO objectives. GRPO introduces a group-level clipping mechanism that allows for more efficient training while maintaining stability.

To summarise, an LLM takes a prompt as an input and predicts the most likely continuation, resultingly "understanding and generating human-like text based on context" [12].

Ding et al [9] dive deep into the uses of LLMs in financial markets in : "Large Language Model Agent in Financial Trading - A Survey". The authors highlight that LLMs are theoretically well suited for the role of professional traders due to 'the ability to process large amounts of information quickly and produce insightful summaries'.

Reviewed LLM agents are divided into two main categories: LLMs as Traders and LLMs as 'Alpha Miners'.

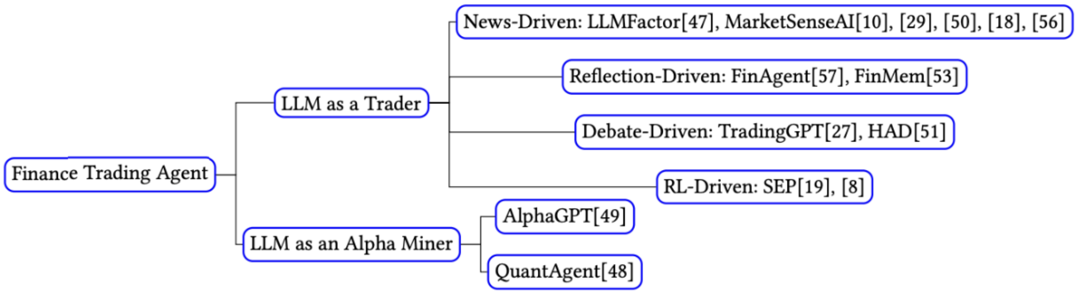

Finance Trading Agent
LLM as a Trader
News-Driven: LLMFactor[47], MarketSenseAI[10], [29], [50], [18], [56]
Reflection-Driven: FinAgent[57], FinMem[53]
Debate-Driven: TradingGPT[27], HAD[51]
RL-Driven: SEP[19], [8]
LLM as an Alpha Miner
AlphaGPT[49]
QuantAgent[48]

Figure 1: Overview of architectures of finance LLM agents.

Fig. 6. Taxonomy of LLM agent types in financial trading, showing the division between LLMs as Traders (directly executing trades) and LLMs as Alpha Miners (providing analysis and insights to support trading decisions). Adapted from Ding et al. [9].

As shown in Figure 6, LLMs can serve different roles in financial trading environments. As Traders, they participate in the market by trading (choosing BUY, HOLD OR SELL) based on analysis and decision-making processes. As Alpha Miners, they extract insights from market data to act as a copilot in identifying profitable opportunities.

Trading LLMs are separated into 4 categories:

- News driven LLMs
- Reflection driven LLMs
- Debate driven LLMs
- RL driven LLMs

News driven LLMs are described as the most fundamental type - with news plugged into prompt context and the LLM instructed to predict stock price movement for the next trading period. Through reviewing many models, more advanced architectures are looked at by the authors which involve 'summary, refinement of news data and reasoning of the relationship between news data and stock price movement'.

Reflection driven models use LLM based summarisation to extract and refine memories. The models progressively distill high level insights from raw observations, employing the reflections to inform trading decisions.

Debate driven LLMs have demonstrated their ability to strengthen reasoning and factual accuracy. Through a hetrogeneous debate framework where specialised LLM agents (e.g, mood agent, rhetoric agent, dependency agent...) debate with each other, improving news sentiment classification performance.

We have already discussed the effectiveness of RL for training LLM outputs with GRPO and deepseek. In the trading context, backtesting with RL is used as an efficient method of generating high quality feedback on trading decisions. Backtesting is clearly an effective way to source rewards in RL.

The authors do note some downsides with LLMs for trading. To begin, for quant trading models numeric data is integral, however LLMs are designed to process text. As a result, numerical data has to be converted to text strings to be fed into a LLM model. However despite LLM's known weaknesses in maths and reasoning, studies have majorly incorporated numerical data.

6 Specific Implementations of LLM Trading Bots

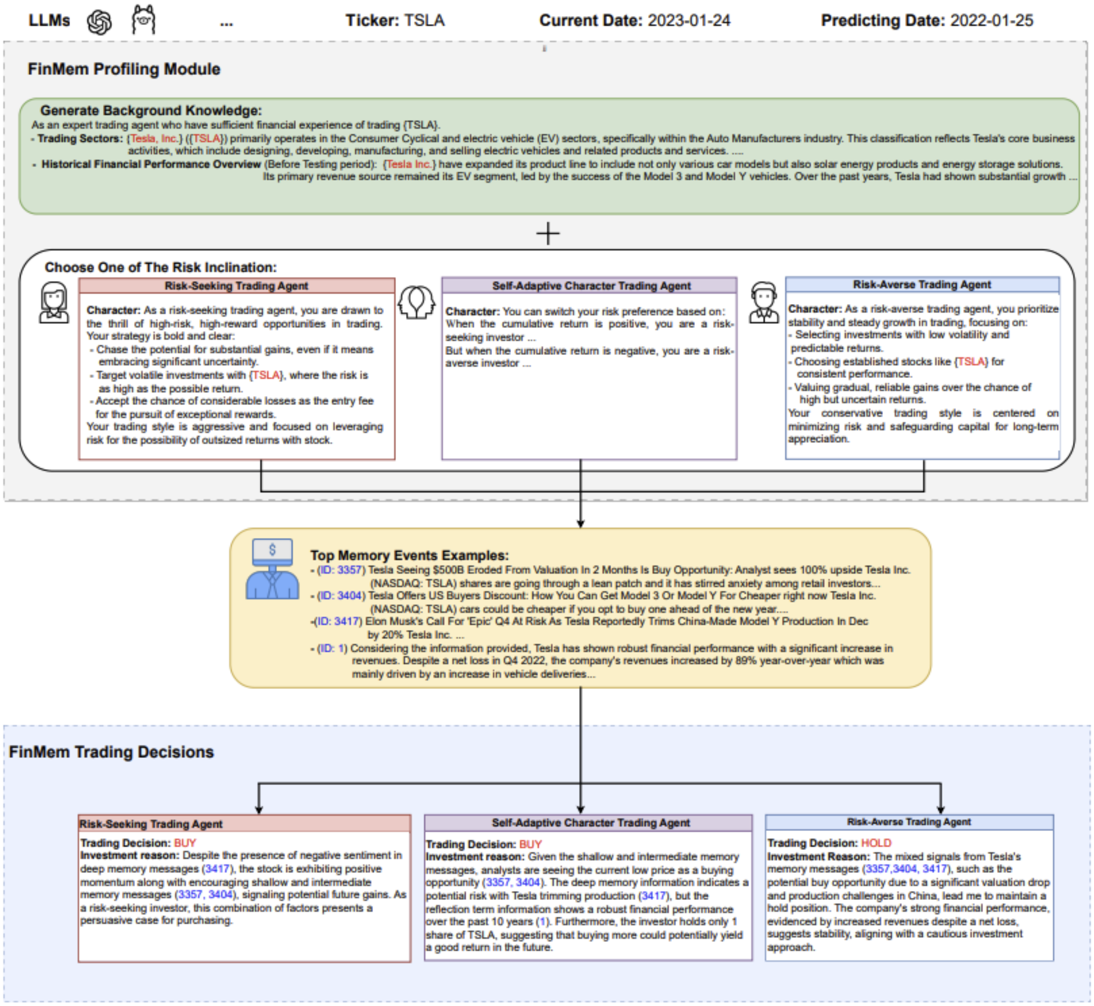
6.1 Trading LLMs

FINMEM: A performance Enhanced LLM Trading Agent with Layered Memory and Character Design Yu et al. [30] at the Stevens Institute of technology have developed FinMem, a "novel LLM Based agent framework for financial decision making". They recognised initially that although LLMs excel at decoding human instructions and processing historical inputs for derive solutions, purpose-driven agents require a supplementary architecture. This is in order to: Process multi-source information, establish reasoning chains and prioritise critical tasks. The FinMEM trading agent is composed of three modules:

[**Profiling ->**] Defines the agent's "character" (ie risk tolerance or trading style)

[**Memory ->**] Layered architecture to process the financial data

[**Decision Making ->**] Transforms insights from memory into long/short/hold actions



Generate Background Knowledge:
As an expert trading agent who have sufficient financial experience of trading {TSLA}.
- **Trading Sectors:** {Tesla, Inc.} ({TSLA}) primarily operates in the Consumer Cyclical and electric vehicle (EV) sectors, specifically within the Auto Manufacturers industry. This classification reflects Tesla's core business activities, which include designing, developing, manufacturing, and selling electric vehicles and related products and services. ....
- **Historical Financial Performance Overview** (Before Testing period): {Tesla Inc.} have expanded its product line to include not only various car models but also solar energy products and energy storage solutions. Its primary revenue source remained its EV segment, led by the success of the Model 3 and Model Y vehicles. Over the past years, Tesla had shown substantial growth ....

Choose One of The Risk Inclination:

Risk-Seeking Trading Agent

Character: As a risk-seeking trading agent, you are drawn to the thrill of high-risk, high-reward opportunities in trading. Your strategy is bold and clear:
- Chase the potential for substantial gains, even if it means embracing significant uncertainty.
- Target volatile investments with {TSLA}, where the risk is as high as the possible return.
- Accept the chance of considerable losses as the entry fee for the pursuit of exceptional rewards.

Your trading style is aggressive and focused on leveraging risk for the possibility of outsized returns with stock.

Self-Adaptive Character Trading Agent

Character: You can switch your risk preference based on: When the cumulative return is positive, you are a risk-seeking investor ...
But when the cumulative return is negative, you are a risk-averse investor ...

Risk-Averse Trading Agent

Character: As a risk-averse trading agent, you prioritize stability and steady growth in trading, focusing on:
- Selecting investments with low volatility and predictable returns.
- Choosing established stocks like {TSLA} for consistent performance.
- Valuing gradual, reliable gains over the chance of high but uncertain returns.

Your conservative trading style is centered on minimizing risk and safeguarding capital for long-term appreciation.

Top Memory Events Examples:
- (ID: 3357) Tesla Seeing $500B Eroded From Valuation In 2 Months Is Buy Opportunity: Analyst sees 100% upside Tesla Inc. (NASDAQ: TSLA) shares are going through a lean patch and it has stirred anxiety among retail investors...
- (ID: 3404) Tesla Offers US Buyers Discount: How You Can Get Model 3 Or Model Y For Cheaper right now Tesla Inc. (NASDAQ: TSLA) cars could be cheaper if you opt to buy one ahead of the new year.....
- (ID: 3417) Elon Musk's Call For 'Epic' Q4 At Risk As Tesla Reportedly Trims China-Made Model Y Production In Dec by 20% Tesla Inc. ...
- (ID: 1) Considering the information provided, Tesla has shown robust financial performance with a significant increase in revenues. Despite a net loss in Q4 2022, the company's revenues increased by 89% year-over-year which was mainly driven by an increase in vehicle deliveries...

Risk-Seeking Trading Agent

Trading Decision: BUY
Investment reason: Despite the presence of negative sentiment in deep memory messages (3417), the stock is exhibiting positive momentum along with encouraging shallow and intermediate memory messages (3357, 3404), signaling potential future gains. As a risk-seeking investor, this combination of factors presents a persuasive case for purchasing.

Self-Adaptive Character Trading Agent

Trading Decision: BUY
Investment reason: Given the shallow and intermediate memory messages, analysts are seeing the current low price as a buying opportunity (3357, 3404). The deep memory information indicates a potential risk with Tesla trimming production (3417), but the reflection term information shows a robust financial performance over the past 10 years (1). Furthermore, the investor holds only 1 share of TSLA, suggesting that buying more could potentially yield a good return in the future.

Risk-Averse Trading Agent

Trading Decision: HOLD
Investment Reason: The mixed signals from Tesla's memory messages (3357,3404, 3417), such as the potential buy opportunity due to a significant valuation drop and production challenges in China, lead me to maintain a hold position. The company's strong financial performance, evidenced by increased revenues despite a net loss, suggests stability, aligning with a cautious investment approach.

Fig. 7. The FinMem architecture showing the three main modules: Profiling, Memory, and Decision Making. Adapted from Yu et al. [30].

Notable is the layered memory module, which was inspired by the cognitive structure of a human trader. It compromises ’sensory’ memory which captures incoming news, reports and prices, ’working’ memory to filter captured inputs and summarise the most salient short-term events and ’long-term’ memory to archive critical signals to be retrieved in later decision cycles.

FinMem was benchmarked against traditional algorithmic strats (ie moving averages and momentum), Deep RL based agents and general LLM agents on large real world stock market datasets. Across many metrics such as cumulative return, Sharpe ratio and Drawdown it outperformed all comparators, demonstrating ability to use external data to profitably trade.

THe model layered memory works using a "top-k" retention parameter which is adjusted in each layer. This is in order to dynamically balance system retention against noise, revealing a clear trade off. If this value is too small, the model misses key signals, reducing returns in high volatility periods. If it is too large however significant noise is introduced which negatively impacts precision in stable markets.

6.2 Alpha Miners

Alpha-GPT: Human-AI Interactive Alpha Mining for Quantitative Investment Wang et al. [29] introduce Alpha-GPT as a human-AI interactive framework for alpha mining by leveraging LLMs through a novel prompt engineering algorithmic workflow. The work is inspired by the fact that traditional alpha mining methods, whether factor synthesising or algo factor mining, have inherent limitations. The goal of Alpha GPT is to provide an interactive and interpretable development loop to faithfully implement quants’ conceptual ideas.

Rather than treating LLMs as black-box predictors, Alpha-GPT uses them as a "thinking core" that generates formulaic alpha expressions from natural language trading ideas which are then refined and evaluated via algorithmic search and human feedback.

Alpha-GPTs workflow consists of four key stages:

Idea Understanding: Users articulate techncial trading concepts, such as "golden cross" "30 day moving average" etc etc, in natural language.

Alpha Generation: The LLM (The paper uses a gpt3.5 variant) produces alpha formulas

Search Enhancement: A genetic programming layer optimises these formulas against a fitness function (Information coefficient). This can double out of sample performance as shown in the below figure.

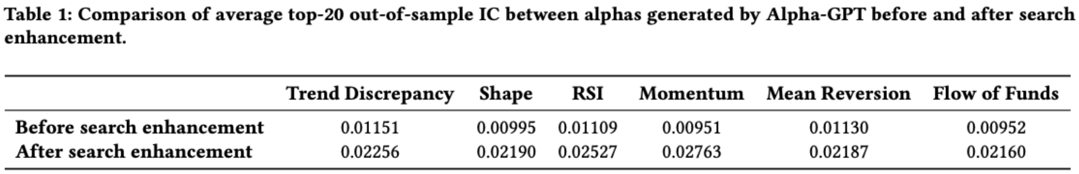
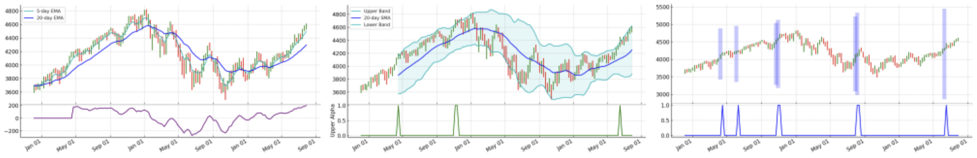


Table 1: Comparison of average top-20 out-of-sample IC between alphas generated by Alpha-GPT before and after search enhancement.

	Trend Discrepancy	Shape	RSI	Momentum	Mean Reversion	Flow of Funds
Before search enhancement	0.01151	0.00995	0.01109	0.00951	0.01130	0.00952
After search enhancement	0.02256	0.02190	0.02527	0.02763	0.02187	0.02160



(a) Golden-cross pattern **(b) Bollinger bands (upper breakout)** **(c) Three white soldiers**

Fig. 8. Performance comparison showing how genetic programming optimization enhances the alpha formulas generated by the LLM. Adapted from Wang et al. [29].

Human AI interaction: Quant researchers finally review, correct and iterate on generated alphas via a chat interface enabling a feedback loop.

6.3 RL Models

A General Framework on Enhancing Portfolio Management with Reinforcement Learning Li et al [20] propose a flexible RL framework to automate portfolio management, constantly reallocating capital across assets to balance return and risk. The framework supports continuous portfolio weights, which enables fractional allocations across a variety of assets across the risk spectrum. In addition, is a long short framework meaning it supports short selling through "tanh" output activation to allow negative weights. Most prior work in this sector overlook many constraints such as tx costs and market liquidity however this paper includes them.

The framework is composed of three distinct policy architectures: Convolutional Neural Network (CNN), Recurrent Neural Network (LSTM) and Multilayer Perceptron for Evolution strategies (ES).

CNN takes a 3d tensor of historical price changes and auxiliary features as state input. LTSM truncates this to a 3 day horizon, whilst ES has a flattened 3-day tensor for all assets.

Experiments span twenty years of market data, with a rolling protocol in which test periods follow training periods in 50 day 10 day ratio. Three RL algorithms: PGAC, PP and ES are benchmarked against rule based baselines like "Follow the Winner" and static equal weight portfolios. Annualised sharpe, max drawdown and cumulative portfolio value are used to evaluate performance.

All RL based strategies except PPO LSTM outperformed the rule based baselines such as "follow the winner". The CNN based PPO agent achieves the highest average portfolio value with lowest drawdown with 1.044x and 0.30 %.

Li et al acknowledge that despite the promising outcomes, training stability remains a concern as high variance suggests the need for more stable training protocols

7 Conclusion

The review synthesises research across the cryptocurrency, RL and LLM domain to chart a pathway towards training LLM based trading agents using GRPO. It has recognised the importance of combining several data sources to get a comprehensive view of market liquidity, participant behavior and sentiment to construct informative state representations. It has explored the capabilities of both LLMs and RL techniques, finally recognising implementation gaps. Buidling on these findings, the thesis will explore building a GRPO trained LLM agent in order to trade cryptocurrency markets.

References

1. 21Shares: Why the Memecoin Mania Isn't a Joke. 21Shares Research (2024), https://www.21shares.com/en-eu/research/why-the-memecoin-mania-isnt-a-joke, accessed: May 1, 2025
2. Almeida, J., Cruz Gonçalves, T.: Cryptocurrency market microstructure: A systematic literature review. Annals of Operations Research 332, 1035–1068 (2024), https://link.springer.com/article/10.1007/s10479-023-05627-5, published Online 27 October 2023; Issue Date January 2024
3. BIS: The crypto ecosystem: key elements and risks . p. 1 (2023)
4. Bloomberg: Trump Memecoin Likely Generated at Least $11 Million in Fees, Analysis Shows. Bloomberg (January 2025), https://www.bloomberg.com/news/articles/2025-01-28/trump-memecoin-likely-generated-at-least-11-million-in-fees-analysis-shows, accessed: May 1, 2025
5. Bloomberg News: Crypto: Citadel securities plans to trade digital coins on exchanges. https://www.bloomberg.com/news/articles/2025-02-24/crypto-citadel-securities-plans-to-trade-digital-coins-on-exchanges (Feb 2025), accessed: 17 April 2025
6. Bouteska, A., Hajek, P., Abedin, M.Z., Dong, Y.: Effect of twitter investor engagement on cryptocurrencies during the covid-19 pandemic. Research in International Business and Finance 64, 101850 (Dec 2022), https://doi.org/10.1016/j.ribaf.2022.101850, published online 20 December 2022; Issue date January 2023
7. Corecco, S., Adorni, G., Gambardella, L.M.: Proximal policy optimization-based reinforcement learning and hybrid approaches to explore the cross array task optimal solution. Machine Learning and Knowledge Extraction 5(4), 1660–1679 (2023), https://www.mdpi.com/2504-4990/5/4/82

8. Demosthenous, G., Georgiou, C., Polydorou, E.: From on-chain to macro: Assessing the importance of data source diversity in cryptocurrency market forecasting. In: VLDB Workshop: Foundations and Applications of Blockchain (FAB), Proceedings of the VLDB Endowment. pp. 1–9 (2024), https://vldb.org/workshops/2024/proceedings/FAB/FAB-6.pdf
9. Ding, H., Li, Y., Wang, J., Chen, H.: Large language model agent in financial trading: A survey (2024), https://arxiv.org/abs/2408.06361, submitted 26 July 2024
10. Felizardo, L.K., Paiva, F.C.L., Costa, A.H.R., Del-Moral-Hernandez, E.: Reinforcement learning applied to trading systems: A survey (2022), https://arxiv.org/abs/2212.06064, submitted on 1 November 2022
11. Fischer, T.G., Krauss, C., Deinert, A.: Statistical arbitrage in cryptocurrency markets. Journal of Risk and Financial Management 14(7), 301 (2021), https://arxiv.org/pdf/2106.00123, accessed: May 1, 2025
12. Gholami, S., Omar, M.: Do generative large language models need billions of parameters? (2023), https://arxiv.org/abs/2309.06589, submitted on 12 September 2023
13. Gladwin, R.S.: What is pump.fun? the solana meme coin factory. Decrypt Media, Inc. (Jan 2025), https://decrypt.co/resources/what-is-pump-fun-the-solana-meme-coin-factory, accessed: 21 April 2025
14. glassnode: An Automated Trading Strategy Grounded in Machine Learning and On-Chain Analytics . p. 1 (2024)
15. Gort, B.J.D., Liu, X., Sun, X., Gao, J., Chen, S., Wang, C.D.: Deep reinforcement learning for cryptocurrency trading: Practical approach to address backtest overfitting (2022), https://arxiv.org/abs/2209.05559, submitted 12 Sep 2022; revised 31 Jan 2023
16. Huang, Y., Zhou, C., Cui, K., Lu, X.: Improving algorithmic trading consistency via human alignment and imitation learning. Expert Systems with Applications 224, 124350 (2024), https://www.sciencedirect.com/science/article/pii/S0957417424012168, published June 2024
17. Jane Street Group, LLC: Jane street. https://janestreet.com (2024), accessed: 17 April 2025
18. Jump Crypto: Jump crypto. https://jumpcrypto.com/ (2024), accessed: 17 April 2025
19. Kalacheva, A., Kuznetsov, P., Vodolazov, I., Yanovich, Y.: Detecting rug pulls in decentralized exchanges: The rise of meme coins. SSRN Electronic Journal (Dec 2024), https://ssrn.com/abstract=4981529, posted: December 4, 2024
20. Li, Y., Wang, J., Cao, Y.: A general framework on enhancing portfolio management with reinforcement learning (2019), https://arxiv.org/abs/1911.11880, submitted 26 November 2019; revised 27 October 2023
21. Nessi, L.: Can ai predict the next 100x memecoin before it pumps? CCN (Feb 2025), https://www.ccn.com/education/crypto/can-ai-predict-100x-memecoins/, accessed: 21 April 2025
22. Qin, Y., Bai, Y., Wang, P., Chen, T., Wang, R., Chen, Y., Liu, H., Wu, Y., Liu, P., Li, G., Zhang, S., Liang, X., Huang, A., Xiao, C., Xu, S., Wang, P., Chu, Z., Xie, Q., Zhao, J., Peng, B., Gu, Y., Zhao, Z., Jiang, D., Liu, M.: Deepseek-v2: A strong, economical, and efficient language model. arXiv preprint arXiv:2402.03300 (2024), https://arxiv.org/abs/2402.03300, accessed: May 1, 2025
23. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms (2017), https://arxiv.org/abs/1707.06347, submitted 20 July 2017; revised 28 August 2017
24. Sharpe, W.F.: Mutual fund performance. The Journal of Business 39(1), 119–138 (1966)
25. Sharpe, W.F.: The sharpe ratio. The Journal of Portfolio Management 21(1), 49–58 (1994)
26. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press, 2nd edn. (2018), https://web.stanford.edu/class/psych209/Readings/SuttonBartoIPRLBook2ndEd.pdf, accessed: 20 April 2025
27. TradingView News: How to find new memecoins before they go viral. TradingView News (2025), https://www.tradingview.com/news/cointelegraph:8d8387a7d094b:0-how-to-find-new-memecoins-before-they-go-viral/, accessed: 21 April 2025
28. Vaswani, .: Attention is all you need (2017), https://arxiv.org/abs/1706.03762, submitted 12 June 2017; revised 2 August 2023
29. Wang, S., Yuan, H., Zhou, L., Ni, L.M., Shum, H.Y., Guo, J.: Alpha-gpt: Human-ai interactive alpha mining for quantitative investment (2023), https://arxiv.org/abs/2308.00016, submitted on 31 July 2023
30. Yu, Y., Li, H., Chen, Z., Jiang, Y., Li, Y., Zhang, D., Liu, R., Suchow, J.W., Khashanah, K.: Finmem: A performance-enhanced llm trading agent with layered memory and character design (2023), https://arxiv.org/abs/2311.13743, submitted 23 November 2023; Revised 3 December 2023
31. Zhao, T., Ding, M., Huang, J., Wen, Z., Huang, S., Yin, Y., Zhao, R., Huang, X., Jiang, J., Niu, Z., Zheng, Y., Luo, Y., Xie, P., Xing, E.P.: Grpo: Generalizable reward with human preference optimization. arXiv preprint arXiv:2406.11903 (2024), https://arxiv.org/html/2406.11903v1, accessed: May 1, 2025