Cover sheet for Research Proposal and Final Dissertation.

Student Name:	Antoine Rithy Chesnay
Student Number:	F118941
Programme Title:	Msc AI and Data Analytics
Dissertation Title:	The role of Deep Reinforcement Learning and Sentiment Analysis in Portfolio Optimisation
Word Count:	9617
Academic Supervisor:	Quan Li
Dissertation route:	Traditional Dissertation

Declaration

By making this submission I confirm that the attached coursework is my own work and that anything taken from or based upon the work of others – or previous work of mine – has its source clearly and explicitly cited; I understand that failure to do so may constitute Academic Misconduct.

I have read the Student Dissertation Handbook and understand the sections on 'Intended Learning Outcomes', 'Academic Integrity', and 'Marking Criteria.'

I have read the Appendix 'Guidance to Students on how to acknowledge, describe and reference the use of Generative AI tools in assessed work' in the Dissertation Handbook. I have acknowledged the use of any Generative AI tools in my dissertation in accordance with this document and understand that a failure to do so will be regarded as academic misconduct by the University.

Tick this box if you do not wish Loughborough University to use your dissertation for administrative, promotional, educational and/or teaching purposes.

Signature: Antoine Chesnay

Date: 17/09/2025

The Role of Deep Reinforcement Learning and Sentiment Analysis in Portfolio Optimisation

Abstract

This dissertation researches the implementation of sentiment analysis with deep reinforcement learning strategies for portfolio optimisation. It addresses a critical gap in quantitative finance and was designed to determine whether sentiment enhanced DRL architectures can outperform traditional DRL architectures. The study uses 10 diversified stocks across different sectors using Proximal Policy Optimization (PPO) and Deep Deterministic Policy Gradient (DDPG) algorithms. The evaluation framework includes a DSR (Differential Sharpe Ratio) reward function, transaction costs, and realistic market constraints.

Both baseline models showed strong performance metrics with DDPG yielding 41.29% total return and 2.80 Sharpe ratio, while PPO generated 35.50% total return and 3.02 Sharpe ratio, both outperforming the S&P 500's 31.49 return. However, contrary to expectations, the integration of sentiment signal caused performance to drop in both the DDPG and PPO architectures. DDPG's performance dropped to 37.85% return and 2.33 Sharpe ratio, and PPO's performance dropped dramatically to 28.24%. This is like due to extreme over conservatism caused by the negative sentiment bias in the sentiment sources.

This research identified key challenges in the integration of sentiment signals, most notably, irregular news coverage, biased sentiment, and PPO's clipping mechanism tending towards conservative strategies. This resulted in PPO demonstrating oversensitivity to sentiment during times of sparse news coverage. Overall, DDPG demonstrated more robustness to sentiment integration in comparison to PPO. The results provide empirical evidence that sentiment integration can degrade results if poorly executed. The key contributions of this paper include providing a comprehensive comparison of DDPG vs PPO for portfolio optimisation, identification of data quality requirements, and a solid framework for realistic evaluation. Further work should focus on developing enhanced sentiment integration techniques from a variety of sources, techniques for debiasing in cases of sentiment imbalances, ensemble approaches which utilise the strengths of different DRL architectures, and dynamic models which adapt sentiment sensitivity depending on market conditions.

While sentiment analysis with deep reinforcement learning for portfolio optimisation is a promising area of research, this dissertation establishes that sophisticated approaches to data quality, feature engineering and DRL design are crucial for the successful implementation of sentiment. The baseline models display strong potential and validate their relevance in quantitative finance; however, sentiment enhancement requires more methodological development than originally anticipated.

Table of Contents

1. Introduction and motivation

As of 2024, the global asset management industry is valued at over $100 trillion. This industry faces challenges in navigating extremely volatile, interconnected, and fluid financial markets. Historically, portfolio management has been shaped by Modern Portfolio Theory (Markowitz, 1952), this is a framework that uses diversification to construct a portfolio which offers the maximum expected return for a given level of risk. While MPT has been critical to the development of modern finance and asset management, its static framework struggles with the increasingly rapid information flow and sentiment shifts that characterise current markets, resulting in outdated risk estimates, inadequate diversification during sentiment-driven crises, and the inability to incorporate real-time market psychology into portfolio construction and rebalancing decisions. The increase of information flow, particularly in asset management and portfolio optimization is a result of the convergence of technological advancement, data proliferation, AI capabilities, regulatory changes, and competitive pressures. The shift from traditional data to alternative data sources has fundamentally transformed how investment decisions are made, with real-time sentiment analysis, social media monitoring, and AI-driven insights becoming standard practice rather than experimental approaches.

It is becoming more apparent that there is a critical need for data-driven portfolio management systems as markets are becoming more driven by market sentiment. The GameStop crash of 2021, which was largely driven by social media sentiment, caused traditional hedge funds to lose in excess of $5 billion because of their failure to account for sentiment-driven market dynamics (Reuters, 2021). Other examples of sentiment driven market crashes include Elon Musk's tweets causing dogecoin's value to swing dramatically, and the 2022 cryptocurrency crash which was partially attributed to social media-driven panic. These phenomena highlight how traditional risk models fail to successfully capture volatility caused by sentiment.

The integration of artificial intelligence into financial markets provides researchers with new academic opportunities. Deep Reinforcement learning (DRL) has proven to be extremely effective in complex decision-making environments (Aggarwal & Kumar, 2022; Benhamou et al., 2024), with studies demonstrating its ability to learn and adapt to changing market conditions while capturing complex patterns in highly nonlinear and non-stationary financial data (Osterrieder, 2023), whereas sentiment analysis is effective at understanding market psychology (Liang et al., 2024) and identifying early warnings of potential risk through the prediction of market volatility and returns (Atkins et al., 2022). However, despite these benefits, there is little existing research of the integration of this in finance, particularly within portfolio optimisation.

This research addresses a critical gap in quantitative finance by investigating how sentiment-enhanced DRL algorithms can improve portfolio performance beyond traditional methods. The findings of this dissertation hold practical relevance for institutional investors, hedge funds, and retail investment platforms, particularly those seeking adaptive, data-driven strategies that leverage both quantitative market signals and qualitative sentiment indicators

2. Research Questions:

This dissertation addresses the following questions:

a. How can deep reinforcement learning algorithms (PPO and DDPG) be designed to effectively integrate sentiment analysis from financial news into dynamic portfolio allocation decisions?
b. What is the incremental value of incorporating FinBERT-derived sentiment signals on the risk-adjusted performance of reinforcement learning-based portfolio optimization systems?
c. How do different DRL architectures (actor-critic vs. deterministic policy gradient methods) perform in sentiment-enhanced portfolio optimization compared to traditional benchmarks?
d. What is the robustness and generalizability of sentiment-enhanced DRL portfolio strategies across different market regimes and time periods?

3. Objectives

Leveraging deep reinforcement learning and natural language processing techniques, this dissertation explores sentiment-enhanced portfolio optimization using financial news analysis and advanced neural network architectures to build an adaptive, sentiment-aware investment framework.

a. Design and implement deep reinforcement learning frameworks (PPO and DDPG) for dynamic portfolio optimization that systematically incorporate sentiment analysis from financial news alongside traditional technical indicators.
b. Develop a comprehensive data integration pipeline that combines multi-modal financial data including price movements, technical indicators, and FinBERT-processed sentiment scores from news articles into coherent state representations for reinforcement learning agents.
c. Create and validate a differential Sharpe ratio-based reward structure that encourages consistent benchmark outperformance while maintaining appropriate risk controls and transaction cost considerations.
d. Conduct rigorous empirical evaluation through temporal train-validation-test splits to assess the performance, robustness, and practical applicability of sentiment-enhanced DRL strategies against traditional portfolio optimization benchmarks across diverse market conditions.

4. Challenges:

A. **Data quality:** Data quality represents the primary challenge for the integration of DRL and sentiment analysis in portfolio optimisation, with fake news and bot-generated content manipulation creating a lot of noise (Fast Company, 2025). In addition to this, low signal-to-noise ratios, survivorship bias, and regime changes can cause difficulty for algorithms designed for stationary environments (Dynamic Datasets FinRL, 2023). Financial time series exhibit heteroskedasticity, low predictability, and large observer effects, features that complicate traditional machine learning algorithms (Maeda et al., 2020). To overcome these challenges, this research uses a Deep Reinforcement Learning (DRL) framework. This is inherently designed to adapt to such non-stationary environments by learning optimal decision-making policies through interacting with the market.

B. **Data acquisition:** A primary limitation encountered during this project was data acquisition. Due to resource constraints, access to premium commercial APIs for historical financial news was not feasible. To circumvent this, an alternative methodology was adopted, leveraging the Internet Archive's Wayback Machine to access and scrape archived news websites which would not typically be accessible. While this strategy successfully collected the desired dataset, it introduced significant operational inefficiencies. The data collection process was severely slowed down by rate limiting protocols which temporarily blocked access if it received a high frequency of requests. This made the web scraping process a tedious task and substantially increasing the data gathering timeline and adding considerable manual overhead to the project.

C. **Computational requirements:** A primary challenge for DRL in finance is the significant computational demands of the models, especially in complex, multi-asset environments. This directly causes a conflict as financial applicants demand low latency. For example, high frequency trading algorithms require sub millisecond response times, however, some transformer techniques used for NLP take longer than this (Wang et al., 2025). As a consequence, live sentiment-based strategies necessitate a trade off between predictive accuracy and high-speed analysis Finance, 2024).

D. **Model drift adaptation:** Continuous retraining is needed for model drift adaptation because financial terminology and sentiment associations change during market stress (DataCamp, 2024). This is particularly relevant to research which includes social media data which is more prone to changing undergo change in environments, however it is also important to handle model drift adaptation when using other sources. Combining different model architectures with different training periods through ensemble methods increases computational complexity, however, will improve robustness across different market regimes (Nexla, 2024).

E. **Overfitting:** Generalisation is very common given the limited training data, this is a critical issue which is also relevant to real-world deployment. Maeda et al. (2020) demonstrate that DRL models can produce predictions that are nonsensical when applied

to out-of-distribution samples. A constant challenge is ensuring model robustness across differing market regimes. This necessitates careful validation procedures to make sure the model is able to generalise successfully past their training environments.

5. Literature review

This section synthesises and evaluates the latest academic developments of portfolio optimisation, deep reinforcement learning, sentiment analysis and their implementation within quantitative finance.

A. Traditional Portfolio Optimisation

Traditionally, Modern Portfolio Theory (MPT), pioneered by Harry Markowitz (1952) in his seminal work, has been central to portfolio allocation strategy, moving investors away from individual analysis of investments to systematic portfolio construction. Markowitz proposed using expected return and variance to measure future returns (1952). The mean-variance optimization framework he developed continues to have relevance in modern day portfolio development and remains mathematically sound. It is focused around minimizing portfolio variance $\left(\sigma^2{}_p\right)$, subject to expected return constraints $\left(E\left(R_p\right)\right)$, and where weights sum to 1. This framework pioneered the revolutionary concept of the efficient frontier and provided a quantitative measure for the benefits of diversification.

Optimisation objective:

$$minimise\ \sigma^2{}_p = \sum_i \sum_j w_i w_j Cov\left(R_i R_j\right)$$

Constraints:

1. Expected return constraint: portfolio expected return must be equal or greater than the target.

$$E\left(R_p\right) = \sum_i w_i E(R_i) \geq E_p$$

2. Full investment Constraint: The sum of the weights of the assets must be equal to 1.

$$\sum_i w_1 = 1$$

The evolution towards equilibrium asset pricing led to the development of the Capital Asset Pricing Model (CAPM), a framework independently established by Sharpe (1964), Lintner

(1965), Treynor (1962), and Mossin (1966). This model extended traditional portfolio theory to encompass market-wide pricing.

The CAPM relationship is defined as:

$$E(R_i) = R_f + \beta_i\big(E(R_m) - R_f\big)$$

This relationship has demonstrated remarkable persistence. Kumar and Aggarwal (2022) documented its continued relevance across 25 subject areas and 128 countries, spanning over six decades of development.

However, whilst MPT and MVO have been instrumental in how investors have designed their investment policies and strategies, these models rely on the assumption of linear relationships, normal distributions of asset returns, rational investors and stationary market conditions. These assumptions limit their applications due to the dynamic and complex nature of real financial markets (Rezaei and Nezamabadi-Pour, 2025).

B. Contemporary developments in Portfolio management

Recent advances in portfolio management have improved traditional approaches to portfolio optimization by resolving some of their key limitations. The Black-Litterman model addresses mean-variance optimisation's high sensitivity to inputs by using market equilibrium returns as a baseline, then uses Bayesian updating to integrate manager views within quantified confidence levels. This is method creates stable portfolios which grow over time rather than unstable and erratic allocations which come from traditional method's reliance on expected return estimates

Risk parity approaches are also gaining more relevance and are becoming increasingly adopted on an institutional level. Adaptive Seriational Risk Parity and Hierarchical Risk Parity using machine learning techniques are addressing the critical assumption that correlations between assets are constant (despite correlations shifting greatly during times of market stress). Both models move beyond the belief that historical correlations determine future correlations and use techniques to determine the underling structural relationship between assets.

Factor investing also embodies another area of evolution within portfolio construction. Factor investing has moved on from the original Fama-French three-factor model (1993) to now a five-factor framework (2015) which now includes profitability and investment factors. This advancement explains the 70-90% of return variation across most markets. Institutional portfolios are now reflecting this shift in methodology by adopting alternative risk premia implementations, this suggests a movement away from market capitalisation weighting towards sophisticated factor-based allocation strategies.

C. Behavioral Finance

Building upon these behavioral insights, the next step in research revolves around investigating how these algorithms can leverage these sentiment-driven opportunities yield higher returns. The field of behavioral finance has established the foundations and the necessary framework for understanding how and why sentiment analysis can improve trading strategies and portfolio optimization.

Behavioral finance emerged in the 1980's as a direct challenge to MPT's assumptions of rational investors. Kahneman and Tversky's (1979) foundational work on Prospect theory highlighted MPT's limitations in describing real-world investor behavior. This theory argues that human decision-making is not always rational, they demonstrated that individuals exhibit loss aversion: feeling the pain of losses approximately 2.25 times more intensely than equivalent gains. MPT treats upside and downside volatility symmetrically, however Prospect theory shows that investors are much more concerned with downside volatility.

Documented anomalies in the market present compelling evidence against traditional models. Research by Jegadeesh and Titman (1993) unveiled a phenomenon known as the momentum effect, their findings show that stocks on winning streaks tend to keep winning, a pattern that shouldn't exist if markets were perfectly efficient. This finding aligns with Mehra and Prescott's (1985) earlier work on the equity premium puzzle. Their research showed that theory historically undervalues stocks, suggesting that behavioural factors influence pricing systematically.

Building upon this empirical work, Behavioural Portfolio Theory was developed to address Modern Portfolio Theory's limitations. Shefrin and Statman's (2000) framework suggested that investors construct layered portfolios, with each layer being influenced by psychological goals rather than directly for portfolio optimisation. This has profound mathematical implications, causing the CAPM two-fund separation theorem to fail to hold when investors exhibit behavioral biases.

D. Sentiment-driven inefficiencies

The Noise Trader Model developed by De Long et al. (1990) and later expanded by Shleifer and Vishny (1997), provides compelling theoretical support for sentiment-based portfolio optimization. This model describes a market environment in which investors are not fully rational, often resulting in poor market timing and exaggerated reactions, or insufficient responses to new information. It also highlights the limits of arbitrage, which allow for persistent inefficiencies in pricing that can be strategically exploited through sentiment analysis.

Building on this foundation, Baker and Wurgler (2006, 2007) introduced an empirical framework to quantify investor sentiment. They developed composite indices incorporating various market indicators, such as closed-end fund discounts, trading volume, dividend

premiums, and IPO activity. Their key insight was that investor sentiment most strongly affects stocks that are particularly hard to value or arbitrage. This finding not only validated the Noise Trader Model but also provided a practical approach for identifying ideal candidates for sentiment-driven strategies: namely, small, volatile, and unprofitable firms with high book-to-market ratios.

Herding behaviour and momentum effects are two key mechanisms which dictate how sentiment affects prices. Scharfstein and Stein's (1990) research reveals that fund managers in some cases engage in herding behaviour to protect their career, while Barberis et al. (1998) show that investors will underreact and overreact to news which can lead to predictable price patterns, which sentiment analysis may be able to systematically exploit.

E. Sentiment Analysis in Finance

The incorporation of sentiment analysis in portfolio optimisation provides every indication that it is one of the most important emerging fields in quantitative finance. Recent evidence shows sentiment-enhanced portfolios achieving annualized returns of 29.3% compared to 19.65% for the S&P 500 benchmark and 17.01% for an equally weighted portfolio, representing outperformance of 9.65 and 12.29 percentage points respectively (Pereira, Caldeira & Sebastião, 2024).

It is becoming increasingly common for major institutional investors to systematically include sentiment factors, with Axioma integrating sentiment-theme factors as standard practice by 2024-2025 (SimCorp, 2025). This shift towards systematic implementation within the industry illustrates the need and creates new opportunities for academic research at the junction of NLP, portfolio optimisation and DRL. As mentioned above, financial sentiment analysis methodology has evolved rapidly, progressing from simple dictionary-based approaches to sophisticated transformer architectures that are able to capture contextual nuances in financial language

Loughran and Mcdonald's (2011) foundational work, demonstrates that general purpose sentiment dictionaries often misclassify financial terms, examples of this would be words like "liability" carrying neutral sentiment in financial contexts despite negative connotations in general usage. This research demanded domain-specific approaches to sentiment analysis which led to the Loughran-McDonald (LM) lexicon being specifically for financial contexts. However, despite these advances in lexicon-based approaches, they still suffer from significant limitations including limited vocabulary, the necessity for manual annotations, and most crucially the inability to capture contextual nuances which are critical for accurate financial sentiment assessment.

The advent of transformer architectures has established a new gold standard for financial sentiment analysis, prominently represented by models like FinBERT (Huang et al., 2023). As demonstrated by Araci (2019) and Huang et al. (2023), adapting BERT's bidirectional architecture through pre-training on specialized corpora, such as Reuters TRC2 and financial

news, yields significant performance gains. When fine-tuned on datasets like the Financial PhraseBank, these specialized models exhibit superior performance, achieving 87% overall accuracy and showing particular strength in identifying negative sentiment with 89.7% accuracy (Du et al., 2024). This capability marks a substantial leap over traditional approaches, which typically perform below 80%. However, while newer models excel on performance metrics, traditional methods still maintain relevance for specific applications. Rule-based systems, such as the FIGAS (Fine-Grained Aspect-based Sentiment) system which uses enhanced lexicons, provide improved interpretability, a crucial advantage in contexts where understanding the model's reasoning is paramount (Loughran & McDonald, 2011).

However, despite this, the key takeback from this is that the performance improvements put transformer models into a different league from traditional classifiers (Shah et al., 2025), underscoring the critical importance of architectural selection for research applications. Even with the more advanced models, feature engineering remains a critical part of the sentiment analysis process. Models must be tailored towards financial-specific terminology ("bear," "bull," "liquid" have different financial meanings), be able to handle numerical expressions (P/E ratios, percentages), and identify temporal context by identifying forward looking statements (Kearney & Liu, 2014). Further development has led to multi-level domain adaptation strategies which are adapted to geographic location, industry, and evolve as market conditions, terminology, and sentiment patterns change over time (Du et al., 2024).

F. Deep Reinforcement Learning in Portfolio Analysis and Finance

Reinforcement learning refers to a type of machine learning where an agent learns by interacting with its environment. The agent receives positive or negative rewards from its actions depending on the reward function and the outcome. This method of learning particularly excels at sequential decision making, especially when the environment is complex and difficult to define.

In recent years, Deep Reinforcement Learning, particularly in the financial context, has garnered considerable attention due to its ability to continuously adapt optimal policies and strategies as market conditions change.

The integration of AI tools represents a paradigmatic shift in quantitative finance. Investors are moving away from traditional methods, focused primarily on Modern Portfolio Theory (Markowitz, 1952) and risk diversification, towards new methods which enable adaptive, data-driven decision-making in complex market environments. Empirical research demonstrates how certain Deep Reinforcement learning strategies, in particular actor-critic methods, such as Proximal Policy Optimization (PPO), Advantage Actor-Critic (A2C), and Deep Deterministic Policy Gradient (DDPG), were able to yield 1.85 times higher annual returns and Sharpe ratios compared to traditional Mean-Variance models based on MPT (Benchmarking RL Algorithms, 2024; Pereira et al., 2024). These results strongly suggest

that traditional models are becoming increasingly outdated for contemporary applications that require dynamic, adaptive strategies.

The institutional adoption of DRL-based advisor systems has caused the convergence of finance and artificial intelligence to accelerate dramatically and highlights the potential in automating investment decisions and portfolios (Huang et al., 2024). While stochastic control approaches (Bellman equations, Optimal control theory, Merton's portfolio problem) rely on model assumptions which are not always accurate, DRL models are not bound to as many assumptions and can leverage large financial data sets while adapting to non-stationarity (Kumar et al., 2022). It is precisely at this intersection of machine learning, behavioural finance, and portfolio theory that my research is situated, offering a novel framework that capitalizes on this shift from rule-based to learning-based models. While theoretically, the advantage of DRL in portfolio optimisation is attractive for investors, the implementation and validation of these methods necessitate careful evaluation and examination. Moving from academic research to institutional applications raises now challenges and limitations which must be addressed carefully

G. Portfolio optimization applications and evidence

Currently, DRL applications in portfolio management have shifted from simple experimental prototypes to more complex and developed framework. DPO (Deep Portfolio Optimisation) framework successfully achieves superior portfolios values and maintains reasonable Sharpe ratios (Yan et al., 2024). Research show that by adopting DRL algorithms risk-adjusted returns improved by 10-25% and Sharpe ratio improved by 0.15-0.4 points, when using stocks from major indices (Risk-Adjusted DRL Portfolio Optimization, 2025). This research validates the utility of DRL beyond simply theoretical advantages.

Feature engineering remains a critical part of the training process, with feature engineering innovations leading to enhancements of performance in portfolio applications. Convolutional neural networks, with longer lookback periods, substantially outperform traditional MLP neural networks for feature extraction due to their proficiency in handling sequential data (MDPI Algorithms, 2024). Integrating technical indicators, historical price data, and volatility metrics produces robust state representations, which enhances the quality of decision-making.

To maintain realistic trading constraints, novel reward functions are used which incorporate transaction costs, risk factors, and market conditions into the Markowitz mean-variance framework (Deep RL for Portfolio Selection, 2024). These risk-cost reward functions allow for environment which more accurately represent the market.

Research into various portfolio rebalancing strategies has provided critical insight, with studies suggesting that periodic rebalancing proves to be more efficient than continuous rebalancing to slippage and transaction costs (MDPI Algorithms, 2024). This suggests that DRL agents should optimise the rebalancing frequency as a part of the learning process.

Evidence of AI being used in portfolio optimisation is not limited to only the US stock market, with studies across Indian (Sensex), US (Dow), Taiwanese (TWSE), and Spanish (IBEX) markets demonstrating consistent DRL superiority. While performance varied, the models were able to outperform their benchmarks (after accounting for risk), across diverse markets (Risk-Adjusted DRL, 2025).

H. Deep Reinforcement Learning Algorithms with Sentiment Analysis Integration for Portfolio optimisation

Among the various reinforcement learning approaches available for portfolio optimization, Proximal Policy Optimization (PPO) has emerged as the leading algorithm for sentiment-enhanced trading strategies. Policy-based algorithms emerge as particularly well suited for continuous action spaces that are typical in portfolio allocation problems, and sentiment integration improves the capability of those algorithms to capture market psychology along with price dynamics. Because empirical studies show that PPO converges more effectively and performs better with risk-adjustment than other algorithms when they incorporate sentiment data (Liu et al., 2024), Proximal Policy Optimization (PPO) has become the dominant choice for sentiment-improved portfolio optimization. The clipped objective function in PPO stops excessive policy updates using noisy financial data. Schulman et al. (2017) note that this addresses a critical challenge within financial applications because market signals and sentiment data contain substantial noise. PPO including sentiment state augmentation achieves 15-20% Sharpe ratio improvements over price-only baselines. This shows a strong level of handling of multimodal data noise in recent implementations like the Multimodal DRL Portfolio (2024).

Building on the success of policy-based methods, actor-critic algorithms represent another promising approach for integrating sentiment data into portfolio decisions. Because their architecture suits frameworks that learn multimodally, actor-critic methods perform well in sentiment-improved portfolio optimization. The Advantage Actor-Critic algorithm does show particularly strong results when it is integrated with sentiment analysis for studies do report it as the top performer in cumulative rewards despite initial expectations favoring SAC and DDPG (Deep Reinforcement Learning Strategies, 2024). A2C updates synchronously thus they provide stability a factor important for financial applications that incorporate sentiment data. In addition to this, A2C can efficiently train on large multimodal datasets that combine price and sentiment features processing in parallel (Mnih et al., 2016). Cross-modal attention mechanisms inside A2C architectures allow for price patterns to influence sentiment processing, or the other way around since they do create advanced sentiment-price fusion strategies (Li et al., 2021).

While on-policy methods like PPO and A2C show strong performance, off-policy algorithms offer distinct advantages for sentiment-enhanced portfolio optimization, particularly in handling complex action spaces. Twin Delayed Deep Deterministic Policy Gradient is quite adept in the handling of continuous action spaces. It also provides sentiment-improved risk management capabilities. Research shows TD3's advantage in creating ideal portfolio plans

within high-dimensional contexts using sentiment data especially when used with mean-variance reward functions joining sentiment alignment with transaction costs plus risk dislike parameters (Deep Reinforcement Learning for Portfolio Selection, 2024). For sentiment integration, TD3's dual critic architecture can prove particularly effective. One critic evaluates price-based Q-values and one processes sentiment-improved Q-values because this architecture takes the minimum for reducing overestimation bias common within multimodal financial applications (Fujimoto et al., 2018). After the delayed policy updates and target policy smoothing mechanisms handle sentiment noise effectively, policy modifications occur when sentiment features stabilize.

As a foundational algorithm in the off-policy category, Deep Deterministic Policy Gradient (DDPG) provides important baseline comparisons for sentiment-enhanced strategies, though it requires more careful tuning than its successors. DDPG offers reliable baselines regarding sentiment-increased portfolio optimization however DDPG needs precise hyperparameter adjustments that weigh price and sentiment signal importance. In comparative studies, sentiment-improved DDPG consistently outperform market benchmarks in annualized returns (Liu et al., 2024), and DQN shows strong performance across different market conditions when researchers augment it with sentiment features (MDPI Algorithms, 2024). However, off-policy algorithms such as DDPG, TD3, also SAC stay vulnerable to noisy rewards within financial environments when sentiment signals conflict against price movements, so on-policy methods prove more reliable through certain sentiment integration applications. Because the DDPG's experience replay mechanism can effectively store sentiment-price-action triplets, it enables efficient learning of historical multimodal data patterns.

Finally, among the maximum entropy approaches to reinforcement learning, Soft Actor-Critic (SAC) presents both opportunities and challenges for sentiment-enhanced portfolio optimization. Entropy is regularized by Soft Actor-Critic (SAC) encouraging exploration of sentiment-price relationships, which benefits volatile market conditions when customary patterns break down. However, the empirical evidence does suggest SAC's performance varies greatly across market regimes when sentiment is integrated. Some studies report how it underperforms in comparison to simpler actor-critic methods throughout stable market periods (Deep RL Strategies in Finance, 2024). SAC formulates maximum entropy offering theoretical advantages for discovering non-obvious sentiment-price relationships. However, financial applications may need researchers to modify it for domains that are specific or integrate it with regime-aware strategies (Li et al., 2021). Since sentiment predictive power does vary over time, temperature tuning automatically in SAC adapts to exploration.

I. Reward function innovation and optimization

Reward function design represents perhaps the most critical innovation area in financial RL applications. Traditional approaches using simple returns prove inadequate for capturing the

risk-return trade-offs central to portfolio management. The Differential Sharpe Ratio (DSR) approach offers several advantages over the traditional Sharpe ratio for reinforcement learning applications due to its enhanced responsiveness and reduced temporal dependencies. Moody & Saffel (2001) were the first to demonstrate that DSR provides more useful and immediate feedback than traditional cumulative portfolio statistics (Sharpe Ratio), which change slowly. This directly addresses issues raised by Deng et al. (2016), who demonstrated that traditional Sharpe ratios have limited action-reward relationships in RL environments due to their reliance on historical portfolio performance. The literature supports DSR's superiority in RL applications with Zhang et al.'s (2019) research demonstrating more stable training and faster convergence rates when using differential metrics compared to cumulative/absolute metrics. In addition to this, Liu et al. (2020) demonstrated that DSR-based reward functions facilitate the development of more robust portfolio strategies that exhibit sustained outperformance across diverse market regimes, thereby establishing DSR as the optimal framework for dynamic portfolio optimization applications.

Building upon the principles that made DSR effective, further innovations continue to refine reward signals for greater nuance and adaptability. Recent advances include Average Sharpe Ratio reward functions, which are specifically engineered for Actor-Critic algorithms to ensure stable convergence while maintaining positive performance (Huang et al., 2024). Another sophisticated approach uses multi-objective reward functions. These combine diverse metrics—such as annualized return, downside risk, and the Treynor ratio—into a modular weighting system that can be customized to different investor preferences while maintaining mathematical tractability.

A more significant methodological leap is the development of Self-Rewarding Deep Reinforcement Learning (SRDRL) (Huang *et al.*, 2023). This paradigm moves beyond static, pre-defined rewards by creating a system that dynamically adjusts its own reward function. By referencing expert-defined metrics like the Sharpe Ratio or Min-Max objectives, the agent learns to refine its goals in response to market conditions. This adaptive approach has demonstrated significant outperformance across major financial metrics, including Cumulative Return, Annualized Return, Sharpe Ratio, and Maximum Drawdown, representing the frontier of intelligent reward design.

6. Methodology

A. Asset Universe

The portfolio consists of 10 carefully selected stocks representing diverse sectors to ensure broad market exposure and reduce overall portfolio risk (aligning with MPT):

- **Technology**: Apple Inc. (AAPL), Amazon.com Inc. (AMZN)
- **Utilities**: NextEra Energy Inc. (NEE)

- **Healthcare**: UnitedHealth Group Inc. (UNH)
- **Consumer Staples**: Procter & Gamble Co. (PG)
- **Industrials**: Caterpillar Inc. (CAT)
- **Materials**: Linde plc (LIN)
- **Real Estate**: American Tower Corp. (AMT)
- **Financial Services**: JPMorgan Chase & Co. (JPM)
- **Energy**: Exxon Mobil Corp. (XOM)

B. Data Period and Splitting

Training Period (January 2015 - December 2017): During this period, the neural network learns from historical patterns, sentiment signals, and technical indicators and trains the DRL agents (PPO and DDPG) to learn optimal portfolio allocation strategies

Validation Period (January 2018 - December 2018): During the validation period, different hyperparameters are tested including learning rates, network architectures, batch sizes, and decay rates to select the best-performing model configurations without contaminating the final test results. This ensures that models generalise beyond training data before the final evaluation

Testing Period (January 2019 - December 2019): During this period, the trained models execute portfolio decisions on completely unseen data to provide final, unbiased evaluation of the DRL agents' performance against benchmark strategies.

The temporal splitting approach is employed to prevent look-ahead bias and maintain realistic trading conditions, ensuring that models cannot access future information when making historical predictions. This mirrors real investment scenarios where portfolio managers take decisions based on historical information without access to future market conditions. The three-year training period provides enough diversity to capture different market conditions (e.g. volatile and stable periods).

C. Data collection and pre-processing pipeline:

The data preprocessing pipeline consists of several stages which need to be conducted sequentially. It is designed to process multi-modal financial data integration while handling data quality and balance issues.

Individual Ticker Processing:

Financial data and technical indicators are collected separately using each the “yfinance” module (unofficial API for yahoo finance).

Financial Data:

Daily stock price data including open, high, low, close, and volume are obtained from reliable financial data providers. All prices are adjusted for splits and dividends to ensure data consistency.

Technical Indicator Calculation:

- **RSI (14-day)**: Relative Strength Index for momentum analysis
- **MACD**: Moving Average Convergence Divergence with standard parameters (12,26,9)
- **Bollinger Bands**: 20-day period with 2 standard deviations
- **Volume Ratios**: Current volume relative to 20-day moving average
- **Price Returns**: 1-day, 5-day, and 20-day returns

Historical News Article Collection: News article headings are webscraped individually for each ticker using a Wayback Machine to access historical versions of trusted financial news sources.

FinBert Implementation: Sentiment analysis is then conducted on the collected news headings using FinBert (Financial Bert).

Sentiment Score Aggregation:

- **Daily averaging**: Multiple articles per day averaged to single sentiment score
- **Confidence weighting**: Scores weighted by FinBERT confidence levels
- **Outlier handling**: Extreme sentiment scores (>2 standard deviations) capped to prevent noise

Individual Ticker Missing Data Processing:

a. **Forward Fill with Exponential Decay:**
b. **Confidence-Weighted Features:** For each ticker, missing data handling creates additional features:

    - **Sentiment confidence score**: Based on recency and article density
    - **Days since last news**: Tracking information staleness
    - **Coverage quality indicator**: Ratio of days with news to total trading days

Data Integration and Normalisation of Features

Individual CSV file: Each the data is integrated into the respective files for each ticker. They will contain:

- **Financial features**: Price, returns, technical indicators
- **Processed sentiment features**: Filled sentiment scores, confidence indicators
- **Metadata features**: Coverage quality, days since news, article density

Normalisation: This ensures that features are all on a similar scale, preventing features with larger magnitudes to dominate features with smaller magnitudes.

1. **Rolling Z-Score (Most Financial Features):** is crucial for financial data because market volatility and price levels change dramatically over time. This approach keeps features normalized relative to recent market conditions rather than historical extremes.
2. **Min-Max Scaling (Technical Indicators**): This is crucial for technical indicators which have natural bounds or meaningful thresholds. This retains the economic interpretability while ensuring that all the features contribute to the neural networks learning.

D. Key components of reinforcement learning:

In recent years, Deep Reinforcement Learning, particularly in the financial context, has garnered considerable attention due to its ability to continuously adapt optimal policies and strategies as market conditions change.

Reinforcement Learning is essentially a machine learning technique where autonomous agents learn to make decisions and evaluate performance by interacting with an environment. The agent continuously updates its policy depending on the reward function, also known as a Markov Decision Process (MDP). In this process, the agent observes a state, takes an action, and receives a reward at each time step.

1. The Agent: This is the decision maker which interacts with the environment.
2. The Environment: The world or context in which the agent operates.
3. State: a snapshot of the environment at a specific time.
4. Action (A): decision made by the agent within a specific state
5. Reward (R): The feedback received from the environment after taking an action.
6. Value Function (V): The future cumulative reward.
7. Policy: The strategy the agent uses.

This approximates the future cumulative reward from taking a specific action. The Bellman equations are a key concept in reinforcement learning which allows the system to evaluate the reward. Mathematically, the Bellman equations are expressed as:

$$V(s) = \max_{a} \left[R(s,a) + \gamma \sum_{s'} P(s'|s,a) V(s') \right]$$

Reinforcement learning loop:

1. Observation
2. Action selection
3. Execution
4. Feedback
5. Learning/Update

6. Repeat

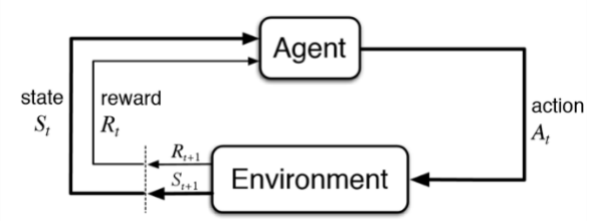

Figure 1: Reinforcement Learning Process (Binay Dhakal, 2023))

Deep reinforcement learning in finance combines the sequential decision-making framework of reinforcement learning with the representational power of deep neural networks to solve complex portfolio optimization problems and extract features. DRL's advantage lies in its ability to identify optimal policies through trial and error within dynamic environments, avoiding the assumptions of static markets which MPT faces (Sutton & Barto, 2018).

This study uses two actor-critic methods used for DRL as the trading agent: PPO and DDPG. The following provides a brief overview of each algorithm.

E. Proximal Policy Optimization (PPO)

PPO is an on-policy learning algorithm which learns a stochastic portfolio allocation policy and aims to enhance the policy gradient by balancing between exploration and exploitation. The PPO framework utilizes two neural networks: an Actor network, which determines the distribution of portfolio weights, and a Critic network, which evaluates the quality of the current market state.

PPO constrains policy updates by using a novel objective function called the "clipped surrogate objective function". This constrains any policy changes in a small range and avoids any destructive large weight updates (Simonini, 2022). The PPO designs and updates the optimal policy by maximising returns.

$$L^{CLIP}(\theta) = \mathbb{E}_t\left[\min\left(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon)\hat{A}_t\right)\right]$$

$r_t(\theta)$: This denotes the probability ratio between the current and old policy.

ϵ: epsilon is the hyperparameter that defines the clip range.

$\hat{A}_t$: This measures how much better (or worse) the action taken was compared to what was expected**.**

E. Deep Deterministic Policy Gradient (DDPG)

Deep Deterministic Policy Gradient (DDPG) is an off-policy reinforcement learning algorithm that effectively learns policies for high-dimensional, continuous action spaces by combining actor-critic methodologies with Q-learning principles. In the DDPG framework, the actor network defines the policy through deterministic mapping of states to specific actions, while the critic network assesses these actions using Q-value estimation. The critic's training process follows the Bellman equation framework:

$$Q(s_t, a_t) = r_t + \gamma Q\big(s_{t+1}, \mu(s_{t+1}|\theta^{\mu})\big)$$

where μ represents the deterministic policy function. The actor network undergoes updates through the sampled policy gradient approach:

$$\nabla_{\theta^{\mu}} J \approx \frac{1}{N} \sum_{t} \nabla_a Q(s, a|\theta^Q) \big|_{s=s_t, a=u(s_t)} \nabla_{\theta^{\mu}} \mu(s|\theta^{\mu}) \big|_{s_t}$$

This dual-network architecture enables DDPG to efficiently manage continuous action spaces, making it particularly well-suited for portfolio optimization where precise, deterministic allocation decisions are required across multiple assets simultaneously.

F. State Space Design

The state space at time t consists of a 72-dimensional vector containing:

For each of the 10 stocks (60 dimensions total):

- 1-day price return
- RSI (14-day)
- MACD signal
- Bollinger Band position
- Volume ratio
- Current sentiment score

$$s_{s,t} = (r_1, RSI_1, MACD_1, BB_1, Vol_1, Sent_1, \dots, r_{10}, RSI_{10}, MACD_{10}, BB_{10}, Vol_{10}, Sent_{10})$$

Portfolio-level features (12 dimensions):

- Current portfolio weights for each stock (10 values)
- Current cash position
- Portfolio volatility (20-day)

$$s_{p,t} = \left(w_1, w_{2,} w_3, \dots w_{10}, cash, \sigma_p\right)$$

G. Action Space Design

The action space for the DRL agent consists of a vector of continuous portfolio weights at time t.

$$a_t = \left(w_{1,t}, w_{2,t}, \dots, w_{10,t}\right)$$

The weights are bound by the following constraints:

a. Long-only constraint

$$w_{i,t} \geq 0 \; \forall \in \{1,2, \dots, 10\}$$

- No short selling allowed
- Each stock weight must be non-negative

b. Full investment constraint

$$\sum (i = 1 \; to \; 10) \; w_i \leq 1$$

- Total allocation cannot exceed 100% of portfolio value
- Cash allowance: Remaining funds $\left(1 - \sum w_{i,t}\right)$ held as cash
- Flexibility: Agent can choose partial investment vs. full investment

c. Maximum Concentration Constraint

$$0 \leq w_i \leq 0.2$$

- Maximum 20% allocation per individual stock
- Risk management: Prevents over-concentration in single assets
- Regulatory compliance: Many funds have concentration limits

The continuous action space allows for flexible portfolio allocation and is particularly

suitable for DDPG, while PPO handles it through appropriate action distribution parameterization.

H. Reward Function Design

The reward function, denoted as R_t, serves to evaluate whether the action of the agent is "good" or "bad". The agent aims to maximise the total reward over time. The reward function in this case employs the Differential Sharpe Ratio (DSR) to provide more responsive feedback and encourage consistent outperformance of benchmark strategies.

Differential Sharpe Ratio Formula:

$$DSR_t = \left(\frac{r_{p,t} - r_{b,t}}{\sigma_{diff,t}}\right)$$

Portfolio return at time t:

$$r_{p,t} = \left(\frac{Portfolio\_Value_t - Portfolio\_Value_{t-1}}{Portfolio\_Value_{t-1}}\right)$$

Benchmark return at time t:

$$r_{b,t} = \frac{\sum_{(i=1\ to\ 10)} r_{i,t}}{10}$$

The risk-free rate (2%) portfolio benchmark is chosen as it represents a naïve but reasonable investment strategy, while using the same assets as the portfolio.

Differential Volatility (Rolling standard deviation of differential returns over 20 trading days):

$$\sigma_{diff,t} = \sqrt{\left(\frac{1}{19}\right) \sum_{j=1\ to\ 20} \left(diff_j - \mu_{diff}\right)^2}$$

Where:

a. Differential return:

$$diff_j = r_{p,t-j+1} - r_{b,t-j+1}$$

b. Mean Differential return:

$$\mu_{diff} = \frac{1}{20} \sum_{j=1}^{20} diff_j$$

I. Transaction Cost Integration

Transaction costs are incorporated into the reward calculation to ensure realistic portfolio performance assessment. When trading, investors face unavoidable transaction costs such as brokerage feeds, bid-ask spreads, and market impact to name a few.

This encourages the DRL agents to develop cost-conscious trading strategies, penalises overtrading, and promotes more stable positions rather than frequent rebalancing. Excluding these costs would result in overestimations of the actual achievable returns and without them,

it would limit our ability to evaluate the applicability of these DRL strategies. In inclusion of transaction costs bridges the gap between theoretical optimisation and practical implementation

Transaction cost calculation and net portfolio return:

The net portfolio return $\left(r_{p,t}^{net}\right)$ is calculated by subtracting transaction costs from the gross portfolio return

$$r_{p,t}^{net} = r_{p,t}^{gross} - TC_t$$

Where:

- $r_{p,t}^{gross}$: Gross portfolio return before transaction costs.
- TC_t: Transaction costs incurred at time *t*.
- $r_{p,t}^{net}$: Net portfolio return after transaction cost (used in DSR calculation)

Transaction costs $(TC_t)$are proportional, set at 0.1% per trade. The cost calculation is as follows:

$$TC_t = 0.001 \times \sum \left|\Delta w_{i,t}\right| \times Portfolio\_Value_t$$

The term $\left|\Delta w_{i,t}\right|$represents the absolute weight change for stock *i.*

J. Evaluation metrics

Risk-Adjusted Return Metrics

1. Sharpe ratio: The Sharpe ratio, is one of the most widely used portfolio evaluation metrics. is used to measure how well the return of an investment compensates for the risk taken. Essentially, it measures the excess return per unit of volatility. In this case, R_p represents portfolio return, R_f represents the risk-free rate and, σ_p is the standard deviation of the portfolio's returns.

$$Sharpe\ Ratio = \frac{R_p - R_f}{\sigma_p}$$

2. Sortino ratio: This is like the Sharpe Ratio, however, rather than penalising all volatility, it will only punish downside deviations of returns. This metric ignores any positive volatility which is beneficial to the portfolio and is often adopted by investors which are primarily concerned about losses.

$$Sortino\ Ratio = \frac{R_p - R_f}{\sigma_d}$$

3. Calmar ratio: This metric is particularly good for evaluating a strategy's resilience and ability to recover from losses. A higher Calmar Ratio indicates that the strategy achieved superior returns while demonstrating resilience to significant drawdowns.

$$\frac{Annualised\ Returns}{Maximum\ Drawdown}$$

Consistency and Stability Metrics

4. **Winning Percentage**: The percentage of trades or periods (e.g., daily, weekly) that are profitable.
5. **Return-on-Investment (ROI)**: This is a straightforward measure of the total return generated by the strategy over a period relative to its cost.

7. Results and Discussion

A. Summary of Key Performance Metrics

Trading Strategy	
Without Sentiment	With Sentiment

Evaluation Metric	DDPG	PPO	DDPG	PPO
Total Return (%)	41.29%	35.50%	37.85%	28.24%
Volatility (%)	15.09%	10.98%	11.69%	29.52%
Sharpe Ratio	2.7998	3.0157	2.33	3.52
Sortino Ratio	4.0582	4.0284	3.0819	14.24
Calmar Ratio	8.7339	6.2852	6.2798	17.24
Max Drawdown	-6.16%	-5.58%	-6.06%	1.67%
Winning Percentage (%)	59.00%	59.51%	60.16%	64.06%
Final Portfolio Value	$14128.77	$13510.27	$13784.60	$12823.95

The evaluation compares four variants: DDPG and PPO without sentiment (baselines) vs. with sentiment. The portfolio value begins at $10,000 and the test period covers 2019.

Table 1: Performance metrics

B. Overall trends

This empirical evaluation identified key patterns and challenged conventional beliefs about sentiment analysis in portfolio optimisation. While both baseline models performed well, the integration of sentiment data had distinctly different effects across both architectures.

The baseline DRL models perfomed well relative to typical MPT benchmarks which typically yield total returns of around 25-30% Sharpe ratios around 1.5-2. The DDPG model with no sentiment yielded the highest total return (41.29%) and portfolio value ($14128.77). Despite having moderate volatility (15.09%) it had a solid Sharpe ratio (2.80) indicating that the model had efficient risk-return trade-offs. The PPO baseline similarly appeared to prioritise stability more than the DDPG which is evidenced by its lower volatility (10.98%) and Sharpe ratio (3.0157), this however coming at the price of a slightly lower total return (35.50%). This can be explained by PPO's clipping mechanisms which avoids any drastic policy changes and leans towards safer, albeit less opportunistic, investment decisions.

Introducing the sentiment data signals extracted by FinBERT provided mixed outcomes which underscore both the benefits and negatives of this method. For the DDPG algorithm, introducing sentiment data lead to a small decline in total return (37.85%) and Sharpe ratio (2.33), however this came with reduced volatility (11.69%) and a higher win rate (60.16%) while the maximum drawdown stayed largely the same. The Sortino (3.0819) and Calmar (6.2798) ratio are also reduced, this alongside the other metrics strongly suggest that the integration of sentiment data reduces DDPG's aggressiveness without significantly improving performance. Possibly due to DDPG not being very sensitive to data, sentiment integration can lead the portfolio to take more of a defensive stance which is better at mitigating risk but will take less opportunities.

The PPO on the other hand, exhibited significantly more drastic changes once sentiment was introduced. Total return fell to (28.24%) and final value ($12,823.95), however the risk

adjusted metrics suspiciously improved: Sharpe ratio (3.52), Sortino (14.24) and Calmar (17.24) ratios, a minimal max drawdown (-1.67%), and the highest winning percentage (64.06%). The risk-adjusted performance metrics for the sentiment enhanced PPO model are inherently flawed and can be attributed due to the imbalanced nature of the financial news data. Preliminary analysis revealed significantly more negative/neutral sentiment scores than positive ones which is likely to cause over conservatism in the PPO agent. As a consequence, the agent acted extremely cautiously and executed significantly fewer trades in comparison to the baseline and had several periods of complete inactivity. This lack of trading activity made calculating standard deviations difficult as the impact of any isolated movements were amplified. This significantly inflated the risk-adjusted performance metrics making them inaccurate and unreliable for comparison and evaluation. The reason why the PPO adopted this over conservatism so much more than the DDPG is because PPO inherently limits policy deviations causing it to stay with conservative behaviour it has learned in training, while DDPG allows for more flexibility to ignore noise and explore.

The contrasting differences between DDPG and PPO and how they handled sentiment data highlight a key challenge in multi model DRL. The architectures displayed different sensitivity to the new features (sentiment), and this divergence in behaviour indicates that careful consideration of the underlying algorithm's learning dynamics and stability characteristics is needed for the successful integration of sentiment.

C. Strengths

Baseline Models: Both DDPG and PPO demonstrated strong performances when using only financial data, with high Sharpe ratios, total returns, and reasonable drawdowns. Both models were able to outperform the S&P 500 with DDPG and PPO achieving (41.29%) and (35.50%) annual returns respectively, while S&P 500 the achieved (31.49%). Thus, proving that DRL can compete well against other models and validating the DRL framework design, reward structure, and data pipeline for price/technical indicators. This aligns with the current academic literature on DRL and its ability to handle nonstationary markets (e.g., Aggarwal & Kumar, 2022).

Volatility: Sentiment integration appears to decrease volatility in models (disregarding the risk adjusted metrics for the sentiment enhanced PPO), leading to more conservative trading strategies which could be more suitable to risk adverse investors. This conclusion agrees with Atkins’s (2022) work on the concept that sentiment can possibly act as a dampener during volatile times.

D. Weaknesses

Underperformance of sentiment models: Contrary to my initial research question, adding sentiment signals resulted in worse results. This could be indicative that both algorithms struggle to handle the higher dimensionality. It could also suggest that sentiment introduces

noise rather than any meaningful signals which leads to the agent learning a suboptimal policy.

Inconsistency across models: While none of the DRL architectures improved once sentiment was introduced, DDPG handled sentiment significantly better. PPO experienced a stark fall in performance suggesting DDPG is potentially more suited for sentiment enhanced DRL strategies.

These findings highlight the broader challenges in sentiment enhancement for finance and illustrate how data heterogeneity can lead to suboptimal performance. The results provide an understanding the incremental value of sentiment and it's ability to enhance downside protection (particularly in PPO due to its over conservatism policy). However, this also reveals limitations in the model's generalisability across different market environments and regimes, especially during times with persistent negative sentiment bias. While deep reinforcement learning architectures display potential, the results highlight the need to implement debiasing techniques to mitigate any imbalances, improve robustness, and increase generalisability.

E. Explanations for underperformance

After having evaluated the performance metrics of the models, it was clear that the models which were supposedly enhanced with sentiment signals were not performing as well as the base models. Analysis of the news data distribution revealed that this was the primary culprit which was responsible for the underperformance of the sentiment-enhanced deep reinforcement learning models in this research. Financial news coverage over specific stocks is typically irregular, with periods of high coverage, followed by periods of low coverage. These gaps in news coverage create significant problems for tracking and modelling market sentiment. Forward filling sentiment in these silent periods can often create an illusion of continuity that doesn't reflect actual market sentiment dynamics. This data quality issue directly impacts the performance of the DRL algorithms. The irregularity of the news flow can cause a signal-to-noise issue where news items during periods of low news flow receive more weight during the sentiment calculations while in times of high news flow, the sentiment processing mechanisms can be overwhelmed. Deep reinforcement learning models require consistent feature representations across time otherwise they struggle to differentiate between genuine shifts in sentiment and periods where sentiment is exaggerated due to data scarcity. This can result in the agent making trading decisions based on noise and not from market sentiment. This subsequently increases portfolio volatility and reduces risk adjusted returns compared to models which rely only on financial data. This raises more questions about the utility of news-based sentiment trading strategies.

Financial markets are incorporating information almost instantaneously, suggesting that by the time data is processed, much of this advantage gained from predictive value will be already included in stock prices as per the Efficient Market Hypothesis (Baker & Wurgler,

2006). Information is becoming increasingly democratised, and consequently, the sentiment derived from news sources may provide less alpha as it would have historically. However, despite this, the evidence from researchers suggests that there are financial benefits from integrating sentiment data however if done incorrectly can lead to diminishing performance. It is possible that the news data acquired misaligns the price actions with sentiment and hence does not obtain any competitive advantage. This requires

Integrating sentiment into the DRL architecture significantly increases the state space dimensionality, making it more prone to overfitting, particularly with limited observations. It is crucial that there is sufficient regularisation of the sentiment features otherwise the model will underperform and while DDPG is good at mitigating this, perhaps more feature engineering is needed on the sentiment data before it can be productively integrated into the DRL algorithms.

Addressing these crucial limitations goes beyond simple data augmentation and requires sophisticated and comprehensive approaches. Most notably, implementing confidence weighting mechanisms which adjust the influence of sentiment based on news density and recency could aid DRL models in adapting to changing data quality and frequency. Sentiment periods of low news coverage could also be augmented by using sentiment sources (e.g. social media, or analyst communications) to generate stronger sentiment signals; however, it is important to note that introducing these new sources can lead to new issues with noise and overfitting to irrelevant features.

The relevance of this research extends beyond the models which were specifically tested. The challenges encountered during this research reflect broader issues in machine learning in finance, most notably, difficulty in extracting meaningful signals from noisy data sources. Furthermore, the temporal mismatch between market reactions and news sentiments causes increased model complexity as price sensitivity to sentiment may vary depending on market conditions.

F. Limitations for industrial implementation

Regulatory and Compliance

1. **Changing Regulatory frameworks:** Regulatory frameworks are constantly changing, as evidenced by new FINRA requirements mandating registration for those

conducting algorithmic trading strategy development (uTrade Algos, 2024). The changing regulations make deployment and data collection extremely difficult.

2. **Black Box Problem**: Regulators will require explanations behind investments, however DRL models provide limited interpretability
3. **Audit trail requirements**: Investment committees need documented justification for changes in portfolio which DRL also cannot easily give.

Data and Infrastructure Constraints

4. **Computational requirements:** Real-time processing of news requires significant computer power and will be costly.
5. **Data Quality Issues**: News coverage varies across different stocks
6. **Latency Problems**: Passing news through FinBERT and a data preprocessing pipeline and updating portfolio weights and allocations will take time. There is a chance that during this time the market conditions can change very quickly.
7. **Storage and Data acquisition Costs**: Maintaining historical news archives requires substantial data storage infrastructure. While cloud can be a solution for data storage, data privacy as well as other needs require on premise solutions for data storage. Substantial operational investment is needed for 24/7 real time processing (Shah et al., 2025) and it is important to note that API costs will be significant.

Market Reality Constraints

8. **Market Impact**: Large changes in the portfolio may have a market impact and affect price. Less liquid stocks (such as LIN and AMT in this experiment) are particularly prone to this.
9. **Market efficiency implications:** Market efficiency implications can create counter intuitive dynamics the more sentiment analysis becomes widespread. While there is evidence that trader and retail sentiment can provide opportunities to gain a profit, Tetlock (2007) poses that adoption inevitably decreases any advantages through arbitrage. This suggests there is value in developing differentiated methodologies.
10. **Capacity Limitations**: Strategies requiring daily news-based rebalancing may not be able to scale beyond a certain point without having a significant market impact
11. **Slippage**: This refers to when actual execution prices differ from theoretical prices used in testing which can impact model performance during back testing.

Technical Limitations

12. **Model Drift**: FinBERT may be trained on historical data and may not have the ability to understand changing financial terms.

13. **Overfitting Risk**: It is possible that the training period does not represent all types of market regimes which can lead to poor performance when testing in unprecedented conditions.
14. **Missing Data Handling**: Forward filling decaying sentiment signals may not accurately represent sentiment.

G. Improvements and Areas for further research

Enhanced data integration remains the most important area for advancement to bridge the gap between theoretical research and practical implementation. It necessitates more diverse and consistent sources of information beyond traditional news articles. It could also prove fruitful to include SEC filings for sentiment analysis to extract forward looking sentiment which is not is not accessible from typical news sources. To capture more retailer investment sentiment, social media platforms such as X and Reddit can be processed using sentiment analysis models specifically designed for finance and social media (FinTweet-BERT). The different sources reflect different market participants sentiment and create a multi-dimensional view which provides a better representation of the market psychology than using only one source for sentiment. To successfully capture the complete psychological underpinnings of financial markets it is essential to integrate more behavioural factors than just sentiment into the model. Using regime detection to identify market phases (e.g. bull, bear, or volatile period) and consequently adjusting sentiment sensitivity would also improve model robustness.

Using more advanced model architectures is the nest step in improving DRL agents' adaptability. While the prototype portfolio in the experiment only used 10 tickers, naturally using more stocks will improve performance and allow for different data integration and handling strategies. Using DRL architectures which implement aggregate sector level sentiment would allow the agent to understand the influence over entire sectors and possibly improve model performance. It is also worth exploring ensemble methods to address the inherent strengths and limitations of each model. For example, during times of high volatility, the ensemble system might favour SAC (Soft Actor-Critic) due to its robust exploration strategy, whereas in stable trending markets DDPG may be preferred. Ensemble approaches to DRL in portfolio optimisation improve robustness by addressing the limitations of each algorithm.

This research needs more comprehensive and rigorous testing on the methodology to ensure robustness across different market regimes. A Walk-forward analysis is recommended where instead of the model being trained once on data, the model is continuously retrained while testing on unseen future data. This approach needs more testing against different market environments, volatilities, and economic cycles to reveal whether the agent's success stems from favourable market conditions or from actual skill. Sensitivity testing is also required to check against overfitting and reveal whether the model success stays consistent across a reasonable range. This would help also help identify any potential improvements and any irrelevant factors.

The areas which offer the most potential for advancement are improving data quality by including a variety of source for sentiment (e.g. social media, SEC filings), creating adaptive models which respond to changing market conditions, and developing the infrastructure to move from academic research to institutional-grade portfolio management systems.

8. Conclusion

This dissertation explored the use of sentiment enhanced Deep-Reinforcement learning strategies in dynamic portfolio optimisation, and empirically evaluates the role of sentiment analysis and deep reinforcement learning in finance

A. Research questions

Research Q1: This research successfully designed and implemented sentiment enhanced DRL frameworks using DDPG and PPO architectures. The state space effectively modelled and stored the financial indicators and FinBERT processed sentiment scores, and the action space included realistic real-world constraints such as concentration limits and transaction costs. However, despite this, neither PPO nor DDPG performed well once enhanced with sentiment signals. Both architectures handled the multimodal data integration poorly and led to worse performance metrics.

Research Q2: Contrary to my expectations, the integration of sentiment signals yielded lower risk-adjusted performances and total return. DDPG saw a fall from 41.29% to 37.85% total return, while PPO experienced a larger fall from 35.50% to 28.24%. Despite this, it appears that integrating sentiment did provide some downside protection as PPO's maximum drawdown fell from -5.58% to -1.67%, however this is likely due to the sentiment imbalance guiding the agents towards over conservatism. These results indicate that the introduction of sentiment generated more noise that predictive power to the DRL algorithms.

Research Q3: The comparative analysis of the DDPG and PPO architectures when enhanced with sentiment signals revealed key differences between the two models. DDPG demonstrated a strong and robust performance and while the performance dipped once enhanced with sentiment, the performance metrics were relatively stable. PPO on the other hand displayed over conservatism once enhanced which led to inflated risk-adjusted metrics and limited trading activity. This suggests that perhaps DDPG is more suited for sentiment enhanced DRL, particularly in periods of negative sentiment. Both baseline models performed well with DDPG leaning towards a more aggressive trading strategy, which had a higher total return than PPO at the cost of higher volatility.

Research Q4: The DRL strategies enhanced with sentiment data display limited robustness across the testing period (2019), this poor performance can be explained by the data quality, particularly the sentiment data imbalances and poor news coverage. The models become

extremely sensitive to sentiment in periods of low news flow which made them unreliable and simply introduced noise. This illustrates the difficulties of integrating, scaling, and deploying sentiment-based DRL strategies in real market conditions.

B. Key findings and contributions.

This research provides concrete evidence that the integration of news/sentiment signals (if done incorrectly) into DRL portfolio optimisation can be counterproductive and can lead to degrading model performance. Despite developing a comprehensive data preprocessing pipeline, this model was still not able to overcome the fundamental data quality issues. The negative sentiment bias created systematic over conservatism issues which significantly affected PPO's performance. While DDPG's performance did suffer, the performance metrics were relatively stable. Implementing a differential Sharpe ratio also proved to be effective, particularly with the baseline models with both outperforming the S&P 500. The integration of transaction cost provided realistic limitations which the DRL architectures learned to work around.

These findings challenge the notion that using sentiment analysis within portfolio optimisation will always improve portfolio robustness and performance. While it doesn't discredit the work of researchers implementing sentiment analysis in portfolio optimisation, it suggests that investors should be careful when integrating sentiment into their trading strategies. Neglecting to address fundamental data quality issues and integration challenges will result in decreased performance.

C. Final Remarks

This dissertation highlights that deep reinforcement learning strategies can be used successfully for portfolio optimisation and shows, however the integration of sentiment analysis demands a significantly more complex approach and increases the complexity greatly. Successful integration of financial sentiment requires addressing challenges in data preprocessing, feature engineering, and model architecture design. The results of this research underscore the need for more developed and tailored approaches to integrating behavioural finance to successfully capture the dynamics of the financial market. It also provides empirical evidence for other researchers who are considering integrating sentiment-based strategies and establishes a foundation for future work by identifying key limitations and areas of vulnerability. The research indicates that while both didn't perform well when enhanced, DDPG appears to show the most promise for sentiment-based strategies. The results ultimately support the development of deep reinforcement applications in finance and critically highlight the importance of data quality, architecture design, and performance metrics.

9. Reference list:

Aggarwal, S. & Kumar, N. (2022) 'How are reinforcement learning and deep learning algorithms used for big data based decision making in financial industries–A review and research agenda', *International Journal of Information Management Data Insights*, 2(2), p. 100094. doi: 10.1016/j.jjimei.2022.100094.

Araci, D. (2019) 'FinBERT: Financial sentiment analysis with pre-trained language models', *arXiv preprint arXiv:1908.10063*.

Atkins, A., Niranjan, M. & Gerding, E. (2022) 'A sentiment analysis approach to the prediction of market volatility', *Frontiers in Artificial Intelligence*, 5, p. 836809. doi: 10.3389/frai.2022.836809.

Baker, M. and Wurgler, J. (2006) 'Investor sentiment and the cross-section of stock returns', *Journal of Finance*, 61(4), pp. 1645-1680.

Baker, M. and Wurgler, J. (2007) 'Investor sentiment in the stock market', *Journal of Economic Perspectives*, 21(2), pp. 129-151.

Barberis, N., Shleifer, A. and Vishny, R. (1998) 'A model of investor sentiment', *Journal of Financial Economics*, 49(3), pp. 307-343.

Benhamou, E., Saltiel, D., Ungari, S. & Mukhopadhyay, A. (2024) 'Advancing investment frontiers: Industry-grade deep reinforcement learning for portfolio optimization', *arXiv preprint arXiv:2403.07916*.

Binay Dhakal (2023). *MDP (Markov Decision Process) — RL (Reinforcement Learning)*. [online] Medium. Available at: https://medium.com/@binaydhakal35/mdp-markov-decision-process-rl-reinforcement-learning-bc85e5d25031 [Accessed 16 Jun. 2025].

Boston Consulting Group (2025). *Global Asset Management Report 2025: From Recovery to Reinvention*. Boston: Boston Consulting Group. Available at: https://www.bcg.com/publications/2025/reinventing-growth-amid-market-volatility [Accessed 31 July 2025].

DataCamp (2024). *Understanding Data Drift and Model Drift: Drift Detection in Python*. Available at: https://www.datacamp.com/tutorial/understanding-data-drift-model-drift [Accessed 22 July 2025].

De Long, J.B., Shleifer, A., Summers, L.H. and Waldmann, R.J. (1990) 'Noise trader risk in financial markets', *Journal of Political Economy*, 98(4), pp. 703-738.

Deep Reinforcement Learning Strategies (2024). *Deep Reinforcement Learning Strategies in Finance: Insights into Asset Holding, Trading Behavior, and Purchase Diversity*. arXiv. Available at: https://arxiv.org/html/2407.09557v1 [Accessed 23 July 2025].

Deng, Y., Bao, F., Kong, Y., Ren, Z. & Dai, Q. (2016). Deep direct reinforcement learning for financial signal representation and trading. *IEEE Transactions on Neural Networks and Learning Systems*, 28(3), pp. 653-664.

Du, X., Xing, L., Mao, K. & Cambria, E. (2024). Financial sentiment analysis: Techniques and applications. *ACM Computing Surveys*, 57(1), 1-38.

Du, X., Xing, L., Mao, K. and Cambria, E. (2024) 'Financial sentiment analysis: Techniques and applications', *ACM Computing Surveys*, 57(1), pp. 1-38.

Dynamic Datasets FinRL (2023). *Dynamic Datasets and Market Environments for Financial Reinforcement Learning*. Machine Learning, Springer. Available at: https://link.springer.com/article/10.1007/s10994-023-06511-w [Accessed 22 July 2025].

Evolution of RL in Finance (2024). *The Evolution of Reinforcement Learning in Quantitative Finance*. arXiv. Available at: https://arxiv.org/html/2408.10932v1 [Accessed 22 July 2025].

Explainable DRL Portfolio Management (2021). *Explainable deep reinforcement learning for portfolio management*. ACM ICAIF. Available at: https://dl.acm.org/doi/10.1145/3490354.3494415 [Accessed 22 July 2025].

Fama-French Five-Factor Model (2015): Fama, E.F. and French, K.R. (2015) 'A five-factor asset pricing model', *Journal of Financial Economics*, 116(1), pp. 1-22.

Fama-French Three-Factor Model (1993): Fama, E.F. and French, K.R. (1993) 'Common risk factors in the returns on stocks and bonds', *Journal of Financial Economics*, 33(1), pp. 3-56.

Fast Company (2025). *Bot farms invade social media to hijack popular sentiment*. Available at: https://www.fastcompany.com/91321143/bot-farms-social-media-manipulation [Accessed 22 July 2025].

Fujimoto, S., van Hoof, H. and Meger, D. (2018) 'Addressing Function Approximation Error in Actor-Critic Methods', *Proceedings of the 35th International Conference on Machine Learning*, Stockholm, Sweden, PMLR 80. Available at: https://arxiv.org/abs/1802.09477
Hachaïchi, Y. and Lanwer, A. (2024) 'Benchmarking Reinforcement Learning (RL) Algorithms for Portfolio Optimization', *ResearchGate*, 11 August. Available at: https://www.researchgate.net/publication/383033429_Benchmarking_Reinforcement_Learning_RL_Algorithms_for_Portfolio_Optimization (Accessed: [date you accessed the paper]).

Huang, A.H., Wang, H., Yang, Y. & Zhao, Q. (2023). FinBERT: A large language model for extracting information from financial text. *Contemporary Accounting Research*, 40(2), 806-841.

Huang, Y., Zhou, C., Zhang, L. and Lu, X. (2023) 'A Self-Rewarding Mechanism in Deep Reinforcement Learning for Trading Strategy Optimization', *Applied Sciences*, 13(2), p. 1121. Available at: https://doi.org/10.3390/app13021121.

Huang, Y.; Wan, X.; Zhang, L.; Lu, X. "A novel deep reinforcement learning framework with BiLSTM-Attention networks for algorithmic trading." Expert Systems with Applications, 2024, 240, 122581.

Kahneman, D. and Tversky, A. (1979) 'Prospect theory: An analysis of decision under risk', *Econometrica*, 47(2), pp. 263-291.

Kearney, C. & Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, 33, 171-185.

Li, Y., Zheng, W. & Zheng, Z. (2021). Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation. *Engineering Applications of Artificial Intelligence*, 97, 104040.

Liang, C., Wei, Y. & Zhang, Y. (2024) 'Financial sentiment analysis: Techniques and applications', *ACM Computing Surveys*, 57(4), pp. 1-38. doi: 10.1145/3649451.

Liang, Z., Chen, H., Zhu, J., Jiang, K. & Li, Y. (2018). Adversarial deep reinforcement learning in portfolio management. *arXiv preprint arXiv:1808.09940*.

Lintner, J. (1965) 'The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets', *The Review of Economics and Statistics*, 47(1), pp. 13-37.

Liu, S. (2024). An Evaluation of DDPG, TD3, SAC, and PPO: Deep Reinforcement Learning Algorithms for Controlling Continuous System. *Atlantis Press*, pp. 15-24.

Liu, X.Y., Yang, H., Chen, Q., Zhang, R., Yang, L., Xiao, B. & Wang, C.D. (2020). FinRL: A deep reinforcement learning library for automated stock trading in quantitative finance. *arXiv preprint arXiv:2011.09607*.

Liu, X.Y., Yang, H., Chen, Q., Zhang, R., Yang, L., Xiao, B. & Wang, C.D. (2019). FinRL: A deep reinforcement learning library for automated stock trading in quantitative finance. *arXiv preprint arXiv:2011.09607*.

Liu, Y., Mikriukov, D., Tjahyadi, O.C., Li, G., Payne, T.R., Yue, Y., Siddique, K. & Man, K.L. (2024). Revolutionising financial portfolio management: The non-stationary transformer's fusion of macroeconomic indicators and sentiment analysis in a deep reinforcement learning framework. *Applied Sciences*, 14(1), 274.

Loughran, T. & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35-65.

Loughran, T. & McDonald, B. (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, 54(4), 1187-1230.

Maeda, I., DeGraw, D., Kitano, M., Matsushima, H., Sakaji, H., Izumi, K. & Kato, A. (2020). Deep reinforcement learning in agent based financial market simulation. *Journal of Risk and Financial Management*, 13(71), 1-17.

Markowitz, H. (1952) 'Portfolio selection', *The Journal of Finance*, 7(1), pp. 77-91. Available at: https://doi.org/10.1111/j.1540-6261.1952.tb01525.x.

MDPI (2021). *Sentiment Analysis for Fake News Detection*. Available at: https://www.mdpi.com/2079-9292/10/11/1348 [Accessed 22 July 2025].

MDPI Algorithms (2024). *A Systematic Approach to Portfolio Optimization: A Comparative Study of Reinforcement Learning Agents, Market Signals, and Investment Horizons*. Available at: https://www.mdpi.com/1999-4893/17/12/570 [Accessed 23 July 2025].

Mnih, V., Badia, A.P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D. & Kavukcuoglu, K. (2016). Asynchronous methods for deep reinforcement learning. *International Conference on Machine Learning*, pp. 1928-1937.

Mohammadshafie, A., Mirzaeinia, A., Jumakhan, H. and Mirzaeinia, A. (2024) 'Deep Reinforcement Learning Strategies in Finance: Insights into Asset Holding, Trading Behavior, and Purchase Diversity', *arXiv preprint* arXiv:2407.09557.

Moody, J. & Saffell, M. (2001). Learning to trade via direct reinforcement. *IEEE Transactions on Neural Networks*, 12(4), pp. 875-889.

Mossin, J. (1966) 'Equilibrium in a capital asset market', *Econometrica*, 34(4), pp. 768-783.

Multimodal DRL Portfolio (2024). *Multimodal Deep Reinforcement Learning for Portfolio Optimization*. arXiv preprint arXiv:2412.17293.

Nexla (2024). *Data Drift in LLMs—Causes, Challenges, and Strategies*. Available at: https://nexla.com/ai-infrastructure/data-drift/ [Accessed 22 July 2025].

Osterrieder, J. (2023) 'A primer on deep reinforcement learning for finance', *SSRN Electronic Journal*. doi: 10.2139/ssrn.4316650.

Pereira, P.L., Caldeira, J. & Sebastião, H. (2024). Artificial intelligence driven portfolio recommendation system: Leveraging sentiment analysis towards improved performance. *Expert Systems with Applications*, 242, 122758.

Reuters Staff (2021). London-based hedge fund that bet against GameStop shuts down - FT. *Reuters*. [online] 22 Jun. Available at: https://www.reuters.com/world/uk/london-based-hedge-fund-that-bet-against-gamestop-shuts-down-ft-2021-06-22/.

Rezaei, M. and Nezamabadi-Pour, H. (2025) 'A taxonomy of literature reviews and experimental study of deep reinforcement learning in portfolio management', *Artificial Intelligence Review*, 58(94).

Risk-Adjusted Deep Reinforcement Learning for Portfolio Optimization: A Multi-reward Approach. (2025) *International Journal of Computational Intelligence Systems*, Available at: https://link.springer.com/article/10.1007/s44196-025-00875-8

Risk-Adjusted DRL (2025). *Risk-Adjusted Deep Reinforcement Learning for Portfolio Optimization: A Multi-reward Approach*. International Journal of Computational Intelligence Systems. Available at: https://link.springer.com/article/10.1007/s44196-025-00875-8 [Accessed 22 July 2025].

Scharfstein, D.S. and Stein, J.C. (1990) 'Herd behavior and investment', *American Economic Review*, 80(3), pp. 465-479.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A. & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.

Shah, S., Krishnan, A. & Patel, R. (2025). Financial sentiment analysis and classification: A comparative study of fine-tuned deep learning models. *Computational Economics*, 63(2), 475-502.

Sharpe, W.F. (1964) 'Capital asset prices: A theory of market equilibrium under conditions of risk', *The Journal of Finance*, 19(3), pp. 425-442.

Shleifer, A. and Vishny, R.W. (1997) 'The limits of arbitrage', *Journal of Finance*, 52(1), pp. 35-55.

SimCorp (2025). *SimCorp launches improved Axioma Worldwide Equity Factor Risk Model*. Available at: https://www.simcorp.com/about-us/news/2025/simcorp-launches-improved-axioma-worldwide-equity-factor-risk-model [Accessed 22 July 2025].

Simonini, T. (2022). *Proximal Policy Optimization (PPO)*. [online] huggingface.co. Available at: https://huggingface.co/blog/deep-rl-ppo.

Sutton, R.S. and Barto, A.G. (2018) *Reinforcement Learning: An Introduction*. 2nd edn. Cambridge, MA: MIT Press.

Tetlock, P.C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3), 1139-1168.

Treynor, J.L. (1962) 'Toward a theory of market value of risky assets', *Asset Pricing and Portfolio Performance: Models, Strategy and Performance Metrics*, pp. 15-22.

uTrade Algos (2024). *Ethical Considerations and Guidelines for Algo Traders in India*. Available at: https://utradealgos.com/blog/ethical-considerations-and-guidelines-for-algo-traders-in-india/ [Accessed 22 July 2025].

Wang, X., Liu, J. & Thompson, R. (2025). Investor sentiment and cross-section of cryptocurrency returns. *Finance Research Letters*, 61, 104923.

Yan, R., Jin, J. & Han, K. (2024). Reinforcement learning for deep portfolio optimization. *Electronic Research Archive*, 32(9), 5176-5200.

Zhang, Z., Zohren, S. & Roberts, S. (2019). Deep reinforcement learning for trading. *The Journal of Financial Data Science*, 2(2), pp. 25-40.

.