Interim Report: Investigating Reinforcement Learning Approaches in Stock Market Trading

Davies Luo

June 5th, 2024

1 Introduction

In recent years, there has been a surge in the development of artificial intelligence (AI) systems, largely fueled by advancements in deep learning technology and the semiconductor industry [1]. The symbiosis of AI and high-performance computing has driven unprecedented growth and innovation in various sectors, including the financial industry [2]. Companies such as Advanced Micro Devices (AMD) and NVIDIA Corporation (NVDA) exemplify this trend, with their significant stock price surge and market share expansion [3].

Financial markets are inherently complex, characterised by high volatility and affected by myriad factors ranging from economic indicators to investor sentiment. Traditional trading and investment approaches often rely solely on human expertise and are limited by the speed and accuracy with which individuals can process information and execute trades [4, 5]. This limitation has spurred interest in leveraging AI, particularly Reinforcement Learning (RL), to develop automated trading systems that can operate at a speed and frequency beyond human capabilities [6].

This dissertation explores the application of Reinforcement Learning (RL) to automate investment strategies within financial markets. It specifically examines the implementation of various RL algorithms, including policy networks and deep Q-networks [7], for optimising stock trading and investment decisions. This study aims to demonstrate how these advanced algorithms can significantly enhance the decision-making capabilities of investors by optimising their strategies in dynamic, real-time market scenarios. By integrating comprehensive financial datasets and economic indicators, this project seeks to develop robust RL models that can adapt to changing market conditions and deliver superior investment performances.

2 Aim and Objectives

2.1 Aim

This dissertation aims to apply Reinforcement Learning (RL) techniques to automate stock trading processes for retail investors, using the semiconductor industry as a case study. By leveraging RL algorithms, the project seeks to develop automated trading strategies that can operate at a speed and frequency beyond human capability and be able to achieve superior outcomes.

2.2 Objectives

This work will focus on three primary objectives:

- Implement RL Algorithms: Implement various RL algorithms, including Policy Networks, Deep Q-Learning, and Convolutional Neural Network-based RL, to develop automated trading strategies [8].
- Integrate Indicators: Integrate financial, economic, and social indicators, such as stock OHLCV data, key economic indices, and Google Trends data, to preprocess comprehensive financial datasets. Use OpenAI Gymnasium to create a simulated trading environment that incorporates these indicators and stock market data.
- Evaluate Performance: Evaluate the performance of RL models on real stock market data, particularly focusing on AMD and Nvidia, and analyse the predictions and returns from these models. Then refine and optimise the RL models to improve their decision-making capabilities and abilities to adapt to dynamic market conditions.

3 Overview of Progress

3.1 Literature Review

A thorough literature review has been conducted, covering essential topics such as reinforcement learning algorithms, financial market analysis, and the integration of related indicators into RL models. Key papers and sources include:

- **Reinforcement Learning Fundamentals:** The study of reinforcement learning (RL) revolves around understanding how agents can learn optimal behaviours through trial-and-error interactions with an environment. This process is mathematically grounded in Markov Decision Processes (MDPs) and the Bellman equation. These foundational concepts are critical as they define the framework within which RL operates.
  - **Markov Decision Processes (MDPs):** An MDP is a mathematical framework for modelling decision-making situations where outcomes are partly random and partly under the control of a decision maker [9]. An MDP is defined by the tuple (S, A, P, R, γ):
    - S: A set of states representing all possible situations in which an agent can find itself.
    - A: A set of actions available to the agent in each state.
    - $P(s'|s, a)$: The transition probability function, which defines the probability of moving from state s to state s' given action a.
    - $R(s, a)$: The reward function, which provides immediate feedback to the agent after taking action a in state s.
    - γ: The discount factor, $0 < \gamma < 1$, which prioritises immediate rewards over future rewards. It ensures that the value of future rewards diminishes over time.

    The objective in an MDP is to find a policy π that maximises the expected cumulative reward over time.

- **Bellman Equation:** The Bellman equation is central to dynamic programming and RL. It provides a recursive decomposition of the value function $V^\pi(s)$, which represents the expected cumulative reward of the following policy π from state s. The Bellman equation is expressed as:

$$V^\pi(s) = E[R(s, \pi(s)) + \gamma V^\pi(s')]$$

This equation breaks down the value of a state into the immediate reward plus the discounted value of the subsequent state, facilitating the computation of the optimal policy by iteratively improving the value estimates [10].

Understanding these mathematical foundations is essential for developing and analysing RL algorithms especially when applied in complex environments such as financial markets.

- **Financial Market Analysis:** Financial market analysis involves exploring various metrics that are crucial for predicting stock prices and making informed trading decisions. Traditional financial metrics such as OHLCV data (Open, High, Low, Close, Volume) provide a detailed view of price movements and trading volumes. These metrics are fundamental for technical analysis, which traders use to identify trends and make buy or sell decisions based on historical price patterns [11].

  Additionally, the number of shares outstanding is a critical metric that impacts market capitalisation and per-share metrics such as earnings per share (EPS). Changes in the number of shares outstanding due to corporate actions like stock splits or dividends need to be adjusted for accurate financial analysis and comparison over time [12].

- **Integration of Financial, Economic, and Social Indicators:** Integrating broader economic and social indicators into financial models enhances the ability to capture market sentiment and macroeconomic conditions. For instance:

  - **S&P 500 and NASDAQ-100:** These indices represent the performance of large-cap U.S. stocks and are often used as benchmarks for the overall market performance. Movements in these indices can influence individual stock prices.
  - **PHLX Semiconductor Index:** This index tracks the performance of companies primarily involved in the semiconductor industry. Given the focus on AMD and Nvidia, this index provides industry-specific insights that are crucial for making informed trading decisions.
  - **Economic Indicators:** Metrics such as the inflation rate, federal reserve interest rate, consumer confidence index, effective federal fund rate, gold price and oil price provide insights into the overall economic environment, affecting investor sentiment and market behaviours.
  - **Google Trends:** This data reflects the volume of searches for specific terms, serving as a proxy for public interest and sentiment towards particular stocks or the market in general. Analysing trends in search volume can help gauge investor sentiment and predict market movements [13].

  Understanding and integrating these diverse indicators are essential for developing RL models that can adapt to changing market conditions and make robust trading decisions.

  The choice of these particular metrics and indices is driven by their proven relevance and utility in financial analysis and trading strategy development. By incorporating a wide range of financial,

economic, and social indicators, the project aims to build comprehensive and adaptive RL models capable of navigating the complexities of the financial markets.

3.2 Data Preparation

As mentioned, the dataset used in this project comprises various sources of financial, economic and social data. These data sources include:

- **Stock OHLCV and Number of Shares Outstanding:** Daily open, high, low, close, and volume data for AMD and Nvidia, these metrics are fundamental in understanding price movements and trading volumes, serving as the basis for technical analysis. Quarterly Number of Shares Outstanding data is crucial for adjusting historical stock prices to reflect corporate actions like stock splits and dividends, to ensure the accuracy of financial analysis and comparability over time [14]
- **Financial, Economic and Social Indicators:** S&P 500, NASDAQ-100, PHLX Semiconductor, Inflation Rate, Federal Reserve Interest Rate, Consumer Confidence Index, Effective Federal Fund Rate, Oil Prices, Gold Prices, and Google Trends. These indicators provide a broader context for market conditions, capturing both macroeconomic factors and investor sentiment.

To create a comprehensive dataset for training the RL models, the following steps were undertaken:

1. **Data Collection:** Data was sourced from reliable financial databases such as Yahoo Finance, Bloomberg Professional Services, Google Trends, and economic reports from the Federal Reserve. Including minute intervals of stock OHLCV Data, daily intervals of financial and economic indicators, weekly intervals of Google Trends data, and quarterly intervals of data in each company's Number of Shares Outstanding.
2. **Data Normalisation:** All longer-period values were resampled into minute intervals, where each minute within the original period was assigned the same value as that of the longer period.
3. **Data Alignment:** All datasets were aligned based on each minute to ensure that each record corresponds to the same time period across different data sources. This involved merging datasets on the date field to create a unified dataframe.
4. **Feature Engineering:** Additional features were engineered to capture more information. For example, moving averages, relative strength index (RSI), and other technical indicators were computed from the OHLCV data.
5. **Final Dataset Creation:** The final dataset was created by combining all the processed features into a single dataframe. This dataframe includes columns for OHLCV data, economic indicators, Google Trends scores, and derived technical indicators.

3.3 Environment Setup and Methodology

The trading environment was developed using OpenAI Gymnasium [15] , tailored to simulate real-world trading scenarios. The implementation and methodology include several key components and methods:

- **Action Space:** Continuous vectors for buy, sell, and hold decisions allow for granular control over trading actions, enabling the agent to adjust the proportion of assets to buy or sell at each step.
- **Observation Space:** A multi-dimensional array incorporating OHLCV data, economic indicators, and investor sentiment metrics. This provides the RL agent with comprehensive market information for informed decision-making [16].
- **Initial Balance:** The agent starts with an initial balance of $1,000,000 to simulate real trading conditions, providing a realistic baseline for evaluating trading strategies.

The methodology for implementing the trading environment includes:

- **Reset Method:** Initialises the environment to its starting state, setting the initial balance, number of shares held, and other relevant parameters, ensuring consistent baselines for reproducible experiments [17].
- **Get State Method:** Retrieves the current state of the environment, including all relevant financial indicators and stock market data, providing the RL agent with a comprehensive view of market conditions [18].
- **Step Method:** Executes the agent's action, updates the environment's state, and calculates the reward based on changes in the agent's net worth. The reward structure incentivises maximising net worth, with the reward function defined as:

$$r_t = \log\left(\frac{NetWorth_t}{NetWorth_{t-1}}\right)$$

  where $NetWorth_t$ is the agent's net worth at time t. This is the logarithmic return of the agent's networth and accounts for the multiplicative nature of returns and stabilises training by penalising large swings in net worth [19].
- **Take Action Method:** Adjusts the agent's balance and shares based on the action taken, calculating the cost or revenue from buying or selling shares and updating the net worth accordingly.

The general approach of training an RL agent proceeds as follows:

- **Simulating an Episode:** At each state, the RL agent determines an action to take based on its policy. The environment then transitions to a new state according to the action taken, following the Markov Decision Process (MDP). This cycle continues for a specified number of time steps, comprising an episode.
- **Updating the Parameters:** After simulating an episode, the objective function $J(\theta)$ is computed as the expected cumulative reward. For policy networks, the objective function is:

$$J(\theta) = E\left[V^{\pi_\theta}(S)\right].$$

  The parameters θ of the RL agent are then updated by computing the gradient of $J(\theta)$ [20].

Initial experiments demonstrated the RL algorithms' ability to make informed trading decisions based on the integrated dataset. However, further optimisation and evaluation are required to enhance model performance and adaptability to dynamic market conditions.

4 Project Plan

The project time frame spans from 22nd April 2024 to the 12th August 2024. For full details, a visual plan is provided in Figure 1. Key milestones include (key activities of each will be illustrated further in the visual plan):

- **Literature Review (22nd April - 12th May):** Completed in the first three weeks. This phase involved extensive research on RL algorithms, financial market dynamics, and the integration of related indicators into RL models.
- **Environment Setup and Data Preparation (13th May - 31th May):** Completed by the end of the second month. This phase included developing the custom trading environment and preparing the dataset.
- **Model Implementation and Testing (1st June - 15th July):** Ongoing, with continuous iterations and improvements. Initial models using Policy Networks and Deep Q-Learning have been implemented and tested within the simulated trading environment.
- **Final Evaluation and Report Writing (16th July - 10th August):** Scheduled for the final weeks before submission. This phase will involve a comprehensive evaluation of the RL models, refining and optimising the algorithms, and compiling the findings into the final dissertation report.

Dissertation

Davies Luo Gantt Chart Timeline Date: 5th June

Topic: Investigating Reinforcement Learning Approaches in Stock Market Trading

Project start: Mon, 4/22/2024

Display week 1

TASK	PROGRESS	START	END
Literature Review			
Research Reinforcement Learning algorithms	100%	4/22/24	4/27/24
Study Markov Decision Processes and the Bellman equation	100%	4/27/24	5/2/24
Explore financial market analysis and Review traditional financial metrics	100%	5/2/24	5/6/24
Investigate integration of related indicators	100%	5/6/24	5/11/24
Research investor sentiment metrics (Google Trends)	100%	5/7/24	5/12/24
Environment Setup and Data Preparation			
Collecting and integrating stock OHLCV data, financial, economic, and social indicators	100%	5/13/24	5/18/24
Summarise findings on RL applications in stock trading	100%	5/15/24	5/20/24
Normalising the dataset for consistency, ensuring data aligned	100%	5/20/24	5/24/24
Developing a custom trading environment using OpenAI Gymnasium	100%	5/24/24	5/28/24
Testing and making sure the environment is rebust	100%	5/28/24	5/31/24
Model Implementation and Testing			
Implementing various RL algorithms such as Policy Networks, Deep Q-Learning, and Convolutional Neural Network-based RL	10%	6/1/24	6/19/24
Conducting initial tests to evaluate model performance and identify areas for improvement.	0%	6/20/24	7/8/24
Iteratively refining the models based on test results and feedback.	0%	6/25/24	7/15/24
Final Evaluation and Report Writing			
Conducting a comprehensive evaluation of the RL models on real stock market data	0%	7/17/24	7/22/24
Refining and optimising the algorithms to improve their decision-making capabilities and adaptability to dynamic market conditions.	0%	7/22/24	7/27/24
Writing and compiling the final dissertation report, including a detailed analysis of the results, discussions, and conclusions.	0%	7/25/24	7/30/24
Preparing supplementary materials such as visual aids, charts, and graphs to support the findings in the report.	0%	7/28/24	8/2/24
Finalising everything for submission	0%	8/3/24	8/10/24

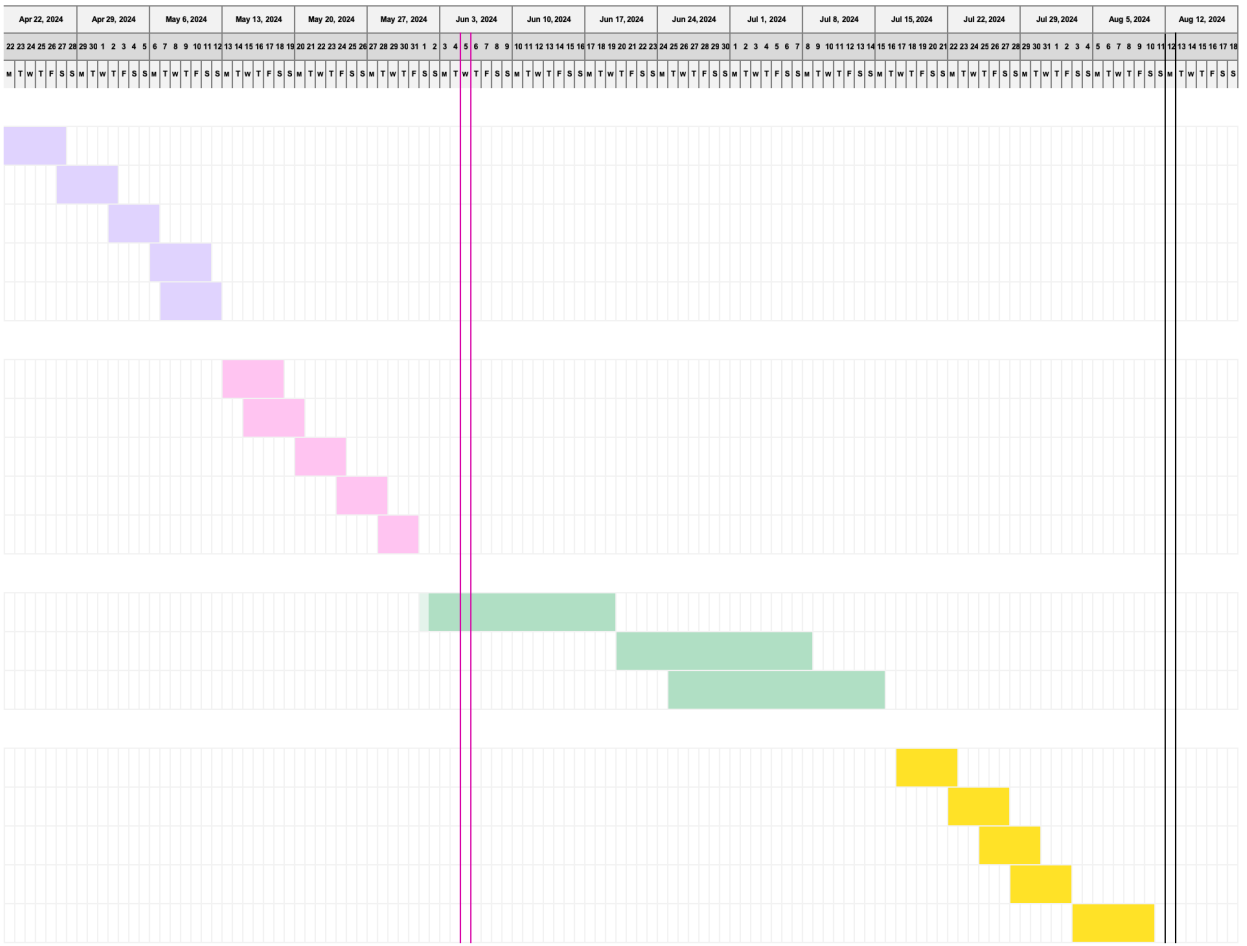

Figure 1: The Visual Plan of the proposed project

References

[1] Batra, Gaurav, Zach Jacobson, Siddarth Madhav, Andrea Queirolo, and Nick Santhanam. "Artificial-intelligence hardware: New opportunities for semiconductor companies." McKinsey and Company, January 2 (2019).

[2] Ahmadi, Sina. "A Comprehensive Study on Integration of Big Data and AI in Financial Industry and its Effect on Present and Future Opportunities." International Journal of Current Science Research and Review 7, no. 01 (2024): 66-74.

[3] Sousa, Bernardo, Cláudia Alves, Marta Mendes, and Manuel Au-Yong-Oliveira. "Competing with Intel and Nvidia: The Revival of Advanced Micro Devices (AMD)." In 2021 16th Iberian Conference on Information Systems and Technologies (CISTI), pp. 1-7. IEEE, 2021.

[4] Arthur, W. Brian. "Complexity in economic and financial markets." Complexity 1, no. 1 (1995): 20-25.

[5] Preda, Alex. "The sociological approach to financial markets." Journal of economic surveys 21, no. 3 (2007): 506-533.

[6] El Hajj, Mohammad, and Jamil Hammoud. "Unveiling the influence of artificial intelligence and machine learning on financial markets: A comprehensive analysis of AI applications in trading, risk management, and financial operations." Journal of Risk and Financial Management 16, no. 10 (2023): 434.

[7] Oh, Junhyuk, Matteo Hessel, Wojciech M. Czarnecki, Zhongwen Xu, Hado P. van Hasselt, Satinder Singh, and David Silver. "Discovering reinforcement learning algorithms." Advances in Neural Information Processing Systems 33 (2020): 1060-1070.

[8] Ansari, Yasmeen, Sadaf Yasmin, Sheneela Naz, Hira Zaffar, Zeeshan Ali, Jihoon Moon, and Seungmin Rho. "A deep reinforcement learning-based decision support system for automated stock market trading." IEEE Access 10 (2022): 127469-127501.

[9] Van Otterlo, Martijn, and Marco Wiering. "Reinforcement learning and markov decision processes." In Reinforcement learning: State-of-the-art, pp. 3-42. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[10] Fei, Yingjie, Zhuoran Yang, Yudong Chen, and Zhaoran Wang. "Exponential bellman equation and improved regret bounds for risk-sensitive reinforcement learning." Advances in neural information processing systems 34 (2021): 20436-20446.

[11] Pahwa, Kunal, and Neha Agarwal. "Stock market analysis using supervised machine learning." In 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), pp. 197-200. IEEE, 2019.

[12] Pontiff, Jeffrey, and Artemiza Woodgate. Shares outstanding and cross-sectional returns. SSRN, 2009.

[13] Petropoulos, Anastasios, Vasileios Siakoulis, Evangelos Stavroulakis, Panagiotis Lazaris, and Nikolaos Vlachogiannakis. "Employing google trends and deep learning in forecasting financial market turbulence." Journal of Behavioral Finance 23, no. 3 (2022): 353-365.

[14] Grinblatt, Mark S., Ronald W. Masulis, and Sheridan Titman. "The valuation effects of stock splits and stock dividends." Journal of financial economics 13, no. 4 (1984): 461-490.

[15] Yue, Naihua, Mauro Caini, Lingling Li, Yang Zhao, and Yu Li. "A comparison of six metamodeling techniques applied to multi building performance vectors prediction on gymnasiums under multiple climate conditions." Applied Energy 332 (2023): 120481.

[16] Kaiser, Gabriele, and Björn Schwarz. "Mathematical modelling as bridge between school and university." ZDM 38 (2006): 196-208.

[17] Pecka, Martin, and Tomas Svoboda. "Safe exploration techniques for reinforcement learning–an overview." In Modelling and Simulation for Autonomous Systems: First International Workshop, MESAS 2014, Rome, Italy, May 5-6, 2014, Revised Selected Papers 1, pp. 357-375. Springer International Publishing, 2014.

[18] Stolle, Martin, and Doina Precup. "Learning options in reinforcement learning." In Abstraction, Reformulation, and Approximation: 5th International Symposium, SARA 2002 Kananaskis, Alberta, Canada August 2–4, 2002 Proceedings 5, pp. 212-223. Springer Berlin Heidelberg, 2002.

[19] De Asis, Kristopher, J. Hernandez-Garcia, G. Holland, and Richard Sutton. "Multi-step reinforcement learning: A unifying algorithm." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, no. 1. 2018.

[20] Wang, Tingwu, and Jimmy Ba. "Exploring model-based planning with policy networks." arXiv preprint arXiv:1906.08649 (2019).

5 Appendix

Data Management Plan

0. Proposal Name

Investigating Reinforcement Learning Approaches in Stock Market Trading

1. Description of the data

1.1 Type of study

This study investigates the application of Reinforcement Learning (RL) techniques to automate stock trading processes for retail investors, using the semiconductor industry as a case study. It involves the development and testing of RL algorithms in a simulated trading environment.

1.2 Types of data

Quantitative data: Financial market data (stock prices, trading volumes), economic indicators, and Google Trends data.
Generated data: Simulation results from RL models.

1.3 Format and scale of the data

The data will be stored in CSV format and managed using Python libraries such as Pandas. The number of records will be in the range of millions for the minute-interval price records of Advanced Micro Devices (AMD) and NVIDIA Corporation (NVDA) and the related financial, economic, and social indicators.

2. Data Collection / Generation

2.1 Methodologies for data collection / generation

The data will be collected from publicly available databases such as Yahoo Finance, Bloomberg Professional Services, Google Trends, and economic reports from the Federal Reserve. Community data standards will be followed to ensure consistency.

2.2 Data quality and standards

Data quality will be controlled through:

- Data validation during collection to ensure accuracy.
- Use of standardized data formats (CSV).
- Regular data audits to identify and correct inconsistencies.
- Peer review of data processing scripts.

3. Data Management, Documentation and Curation

3.1 Managing, storing and curating data

Data will be stored on secure university servers with regular backups. Data will be managed using version control systems like Git to track changes. Community-agreed data standards (e.g., CSV format) will be used for ease of sharing and long-term validity.

3.2 Metadata standards and data documentation

Metadata will include descriptions of data collection methods, data processing scripts, and documentation of variables and records. This information will be stored alongside the data in README files and metadata documentation.

3.3 Data preservation strategy and standards

Data will be preserved on secure university servers for the long term following the procedure. Formal preservation standards, such as those recommended by the university's data management policy, will be followed.

4. Data Security and Confidentiality of Potentially Disclosive Information

4.1 Formal information/data security standards

Since the training data is entirely public data, the formal information/data security standards are not relevant.

4.2 Main risks to data security

Since the training data is entirely public data, security issues with third-party storage solutions (raised by ncl.ac.uk `https://www.ncl.ac.uk/library/academics-and-researchers/research/rdm/working/`) are not a concern. Any trained networks created are intended to support philanthropic causes, encouraging public distribution.

5. Data Sharing and Access

5.1 Suitability for sharing

The data I will use is entirely publicly available and widely peer-reviewed, cited in many published papers. This project aims to eliminate bias in existing datasets, thus encouraging the sharing of code and experiments.

5.2 Discovery by potential users of the research data

All code libraries will be available in the public project GitHub repository (`https://github.com/daviesluo/-Reinforcement-Learning-Approaches-in-Stock-Market-Trading/tree/main`). If the project successfully fulfills its aim, publication will be sought due to its universal application for fair AI across domains, and a DOI will be generated to facilitate discovery.

5.3 Governance of access

As all the data used will be available on a data repository (`https://data.ncl.ac.uk/`), the repository will have control over deciding the access of a potential new user. In addition, the trained networks will be publicly available on GitHub, ensuring transparency and openness.

5.4 The study team's exclusive use of the data

The study team does not have any exclusive data usage. The intent is to share the data, code, and trained models openly with the broader research community. Therefore, we do not have any policy on the exclusive use of the data.

5.5 Restrictions or delays to sharing, with planned actions to limit such restrictions

Proposed data-sharing procedures will be explicitly stated in the repository README. Given the fully public nature of the datasets used, no significant restrictions on data sharing are expected. The intellectual property from developed models will be freely distributed to advance the field, with terms explicitly stated in the project README. The author will not assume any liability for misuse of the developed software. Any delays will be due to data processing and anonymization requirements, which will be minimized through efficient workflows.

5.6 Regulation of responsibilities of users

Since the training data is entirely public data, there is no formal regulation of responsibilities for accessing them. However, the intellectual property from developed models will be freely distributed to advance the field, with terms clearly stated in the project README. The author will not assume any liability for the misuse of the developed software.

6. Responsibilities

The PI is the only one responsible for the study-wide data management, metadata creation, data security, and quality assurance of data.

Resources Required

- Google Colab Pro Plus @ £45.9 / month
- Google Drive Storage (minimum 200GB) @ £2.99 / month

7. Relevant Institutional, Departmental or Study Policies on Data Sharing and Data Security

Policy	**URL or Reference**
Data Management Policy & Procedures	`https://www.ncl.ac.uk/media/wwwnclacuk/research/files/ResearchDataManagementPolicy.pdf`
Institutional Information Policy	`https://services.ncl.ac.uk/itservice/policies/InformationSecurityPolicy-v2_1.pdf`
Other	

8. Author of this Data Management Plan

Davies Luo
+44 7903082497
z.luo24@newcastle.ac.uk