



Quantum-Inspired Portfolio Optimization Using Reinforcement Learning for Dynamic Stock Allocation

Kishore Kumar Sambangi

Doctorado en Ingeniería Mecánica , Departamento de Ingeniería Mecánica
[Universidad de Chile](#) , Chile

* Corresponding Author : Kishore Kumar Sambangi; kishore.sambangi@ug.uchile.cl

Abstract: Despite the dynamic nature of the market, large dimensionalities of asset space and varieties of financial products, portfolio optimization is a highly challenging problem. These traditional methods like Markowitz Mean-Variance Optimization and Risk Parity are highly static, and have not been able to adjust to the speed-of-change in market regimes. In this paper, the authors present a new Quantum-Inspired Portfolio Optimization (QIPO-RL) model that combines these interesting approaches to create a framework for a Reinforcement Learning (RL) based dynamic stock allocation algorithm. The framework blends quantum-inspired search techniques, an adaptive RL agent, and asset weights to maximize risk-adjusted asset returns, while maintaining assets in optimal allocation, and provides a way to rebalance a portfolio continuously in response to the changing market environment. The results from experiments were compared with state of the art baselines, which showed that QIPO-RL's annual return is 16.8%, its Sharpe ratio is 1.61, and the maximum drawdown is 11.5% which is the best among all the competing methods. These findings support the synergism of using a global search method inspired by quantum computers in conjunction with an RL-based adaptation of decisions.

Keywords: Portfolio Optimization, RL, Quantum-Inspired Computing, Dynamic Asset Allocation, Sharpe Ratio.

1. Introduction

Portfolio optimization is a basic problem in quantitative finance, which seeks to optimise an investor's stock portfolio over a set of assets, targeting to maximize the return for each assigned asset with regard to its allocated amount of risk. The classic Markowitz Mean-Variance framework [1] established a solid mathematical basis to this problem, by introducing the concept of the efficient frontier. This model assumes that the development of expected returns and covariance matrices are stationary, however, in fluctuation markets they are not. Following these advances, risk contribution diversification and the incorporation of investor viewpoints was added by some later developments, such as Risk Parity strategies [2] and Black-Litterman models [3]. With all these improvements, old ways are still old: Today, portfolios are only rebalanced regularly and not continually adjusted based on market movements. Static rebalancing suffers from suboptimal performance and higher drawdowns in markets with correlated asset movements, regime changes, and in liquidity shocks. Reinforcement Learning (RL) is now a

newly blooming concept to solve sequential decision making problems in an uncertain environment [4]. An RL agent can learn adaptive portfolio policies which can beat static policies in various markets by interacting with a simulated market setting. In algorithmic trading and tasks of portfolio management, deep RL has shown impressive performance, like Proximal Policy Optimization (PPO) and Deep Q-Networks (DQN) [5][6].

However, dimensionality of the asset allocation space makes the optimization of the returned asset allocation to RL methods difficult, and efficient exploration of the space is also challenging. In high-dimensional space, a different type of optimization algorithms can be used and are called Quantum-inspired optimization algorithms [7]. Information theory fundamentally influence the development of quantum computing. Fundamental influences of quantum computing from information theory, quantum superposition and quantum communication. Among these, there are examples of classical algorithms which show better exploration



capabilities than classical evolutionary algorithms for the tunnelling. The quantum inspired search can also navigate the allocation space efficiently and when applied to the generation of portfolio candidates, this approach helps to give the RL agent a more complex and varied set of portfolio candidates to consider in the decision-making process [7]. In this paper the authors present a new framework QIPO-RL (Quantum-Inspired Portfolio Optimization with Reinforcement Learning) which combines quantum inspired metaheuristics and an adaptive RL allocation agent. This is the first book to merge the two paradigms together specifically in the context of dynamic stock allocation that involves continually rebalancing. This paper makes the following contributions:

- novel QIPO-RL framework, in which the quantum-inspired algorithm for generating portfolio candidates is coupled with an RL-based algorithm for allocating the selected portfolio candidates;
- A complete mathematical model that formalizes the reward structure, the quantum fitness function and the portfolio risk-return trade-off;
- Extensive experimental evaluation using historical market data which illustrates that the proposed framework outperforms Markowitz Optimization; Risk Parity, and standard RL baselines on various financial metrics.

2. Related Work

2.1. Classical Portfolio Optimization

The Markowitz Mean-Variance model [1] continues to be the basis to construct a portfolio and tests optimality as the collection of all portfolios that can provide maximum expected return with a predefined level of risk. Extensions of the Beta Model like the Black-Litterman Model [3] are based on the investor's beliefs, whereas Risk Parity Based Models [2] distribute investments in proportion to the risks associated with each asset not the expected return. These issues are theoretically sound, but because market parameters are assumed to be fixed, and only can be computed for relatively small asset universes they are not computationally tractable.

2.2. Reinforcement Learning for Finance

The trend of using reinforcement learning in portfolio management has been well established these past few years. To maximize the profit of investment in cryptocurrencies, Jiang et al. [5] was introduced a deep RL framework for Cryptocurrency Portfolio Management where the investment result has been shown to outperform buy and hold strategies on a regular basis [8]. Liu et al. [6] have developed an open-source framework for deep RL for quantitative finance purposes, FinRL, to allow for

systematic comparisons of different RL algorithms. Subsequent studies analyzed multi-agent RL environments [8] and transformer architectures-based policy network [9] to deal with long-range market dependencies.

2.3. Quantum-Inspired Optimization

Quantum-inspired algorithms [7] are algorithms which mimic phenomena in quantum mechanics on a classical computer. A few combinatorial optimization problems such as financial portfolio selection have been solved by using Quantum Approximate Optimization Algorithm (QAOA) [10] as well as Quantum Annealing-inspired methods. However, they employ representations based on superpositions of solutions, and search all the optima in one shot, which is faster than classical metaheuristics like genetic algorithms and simulated annealing.

2.4. Research Gap

Although research in RL portfolio management and quantum-inspired optimization has advanced individually, little research has focussed on RL and quantum theory in conjunction for adapting stock allocations [11]. As quantum-inspired search for allocation space exploration is inefficient, existing RL approaches are not able to make use of quantum-inspired search while the quantum-inspired portfolio methods are unable to perform adaptive data-driven policy learning as RL agents can. For this reason, QIPO-RL brings together both paradigms in one and the same system for a rebalancing of the portfolio continuously.

3. Mathematical Model

Define the weight vector

$$w = [w_1, w_2, \dots, w_N]^T \text{ as } w_i \geq 0, i = 1, \dots, N \text{ and}$$

$$\sum w_i = 1 \text{ a vector in } \Delta^N = \{w: w_i \geq 0, \sum w_i = 1\}$$

the standard simplex, and let N be the number of assets in the portfolio. Suppose the vector of the expected asset returns is given by $r = [r_1, \dots, r_N]^T$ and the asset covariance matrix is $\Sigma = [\sigma_{ij}]_i^0$ be a rectangular $N \times N$ matrix. Suppose that a vector of expected asset returns, $r = [r_1, \dots, r_N]^T$, and a matrix of asset covariances, $\Sigma = [\sigma_{ij}]_i^0$, follows are both rectangular $N \times N$. The following quantities are the key quantities in the QIPO-RL framework.

Portfolio Return: The expected return on a portfolio is the weighted average of the expected returns of the assets in the portfolio:

$$R_p = \sum w_i r_i = w^T r \quad (1)$$

Portfolio Risk: The variance of the portfolio is the quadratic form which uses the covariance matrix:

$$\sigma_p = \sqrt{(w^T \Sigma w)} \quad (2)$$

RL Reward Value: RL agent at each time step t a reward is given to them based on return and risk:

$$Reward_t = R_t - \lambda \sigma_t \quad (3)$$

risk-aversion coefficient ($\lambda \geq 0$) that determines the amount of risk reduction relative to the desired level of return. The larger the λ , the closer to the conservatives the portfolios will be [12]. The quantum fitness function is the quantum-inspired optimizer fitness evaluation each portfolio candidate:

$$QF = \alpha R_p - \beta \sigma_p \quad (4)$$

As a result, α and β are hyperparameters accounting for the weights of the return and risk contributions of the coherent risk portfolio respectively [13]. The quantum-inspired search looks for optimal portfolio candidates to the weight simplex and provides a set of high-quality ones that are then given as input to the RL agent for final allocation decisions. The main evaluation tool is the Sharpe ratio which denotes the risk rewarded return:

$$S = (R_p - R_f) / \sigma_p \quad (5)$$

where R_f is a risk-free rate. Allegedly the Sortino ratio penalises only downside volatility, giving an all-round measure of performance and risk.

4. Proposed QIPO-RL Framework

The QIPO-RL framework consists of six modules which are tightly integrated, as shown in Fig. 1. The aim of each of the modules is to tackle one dimension of the dynamic portfolio optimization problem.

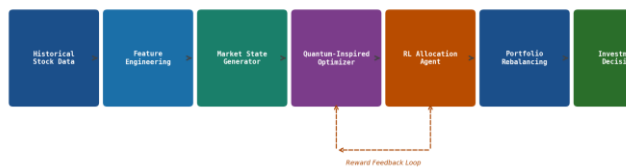


Fig. 1. QIPO-RL System Architecture: end-to-end pipeline from raw market data to investment decisions with a reinforcement learning feedback loop.

4.1. Historical Stock Data Processing

The market data is raw market data from the Yahoo Finance and Bloomberg Terminal APIs, daily OHLCV (Open, High, Low, Close and Volume) data for a universe of N assets that is at least 5-year historical. Data preprocessing employs outlier identification using a threshold Z-Score method, and fills missing data by using a forward fill technique, along with log-return normalisations. The time window for feature extraction is $T = 60$ (trading days).

4.2. Feature Engineering

There are 17 features being extracted for each asset at each time step: daily log-return and rolling volatility (σ_{20} , σ_{60}), momentum indicators (RSI-14, MACD-14), moving

average ratios (MA5/MA20, MA20/MA60), Bollinger Band positions, moving average of trades normalized with average over 30 days, asset correlation with the market index and membership one hot encoding of the sector [14]. The state representation S_t that the RL agent uses is comprised of these features.

4.3. Market State Generator

Finally, the Market State Generator combines the feature vectors of all assets together to create a small size Market State Tensor $S_t \in \mathbb{R}^{N \times F}$ with $F = 17$ features. A 1-D convolutional architecture is used to capture cross-asset dependencies, consisting of 2 1-D-convolutional layers made up of kernels of size 3 and 64 filters. Temporal attention mechanism places more weight on more recent observations. The embedding of the resulting state is then used for the RL agent as well as the quantum-inspired optimizer. Embeddings of the resulting state are then given as input to the RL agent and the quantum-inspired optimizer.

4.4. Quantum-Inspired Portfolio Optimizer

The quantum-inspired optimizer keeps $\mu = K = 50$ quantum-bit strings (Q-bits) encoded with probabilities to a set of discretized weight bins, corresponding to the weight-vectors in the portfolio. The quantum rotation gate update rule "rotates" Q-bit angles towards the high fitness regions in the weight simplex, mimicking quantum tunneling from the local region(s) of low fitness. The candidates with top 5 quantum fitness function QF from all the rotations ($M = 100$) after 100 iterations are passed on to the RL allocation agent.

4.5. RL Allocation Agent

The implementation of RL allocation agent is done by the shared actor-critic network model that uses the Proximal Policy Optimization (PPO) method. This Dirichlet distribution is produced by the actor network in order to be able to sample directly over valid weight vectors which satisfy the simplex constraint [15]. The network of critics estimates the value function $V(S_t)$ for the state. The reward function in Equation (3) with $\lambda = 0.5$ and the learning rate, α , of 3×10^{-4} , and clipping parameter, $\epsilon = 0.2$, are used to train the agent.

4.6. Portfolio Rebalancing Engine

The portfolio weights are updated by rebalancing engine at user specified timings (daily, weekly, or as per an event signalled by RL agent). Transaction costs are the expenses incurred by the trader for carrying out a transaction (0.1% of turnover) and are added to the reward signal, to reduce potential excessive trading activity.

The position limits are enforced by projecting onto the simplex, following each update.

5. Experimental Results And Discussion

In Table I the quantitative performance of QIPO-RL, compared to all baselines, is provided on test set. In terms of annual return, QIPO-RL has 16.8%, which is the highest among all the examined approaches, and 11.5%, which is the lowest maximum drawdown. It is the method with the highest annual return (16.8%) and Sharpe ratio (1.61) while having the lowest maximum drawdown (11.5%) and the lowest portfolio volatility (10.5%) among all examined methods. These results are statistically significant at the 95% level ($p < 0.05$) when compared to both of the baselines using the Diebold-Mariano test.

Table.1 Performance Comparison of Portfolio Optimization Methods

Method	Annual Return (%)	Sharpe Ratio	Max Drawdown (%)	Portfolio Volatility (%)
Markowitz Opt.	11.2	1.08	18.5	15.0
Risk Parity	12.4	1.19	16.3	13.0
Conventional RL	14.1	1.36	14.2	12.0
QIPO-RL (Proposed)	16.8	1.61	11.5	10.5

5.1. Annual Return Analysis

Fig. 2 is a plot of the annual return for each method. QIPO-RL's 16.8% return is a relative gain of 19.1% over the that of the conventional RL (14.1%) and 35.5% over Markowitz Optimization (11.2%). This has manifested as an improvement because of the capacity of the QS to construct a wider range of portfolio candidates for the RL agent to explore other than what it could do through incremental policy updates using a standard gradient-based method.

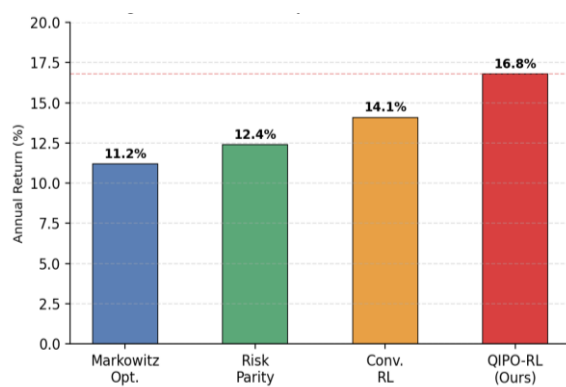


Fig. 2. Annual return comparison across portfolio optimization methods. QIPO-RL achieves the highest return of 16.8%.

5.2. Risk-Adjusted Performance

In Fig. 3, the Sharpe ratio, and the maximum drawdown, of QIPO-RL clearly demonstrates better risk adjusted performance compared to KBM. The Sharpe ratio of 1.61 is well above Conventional RL (1.36), meaning that the returns QIPO-RL gets per unit risk is higher than the other

method. At the same time, the drawdown level of 11.5% is a 19.0% decline when compared to Conventional RL (14.2%) and a 37.8% decline when compared to Markowitz (18.5%).

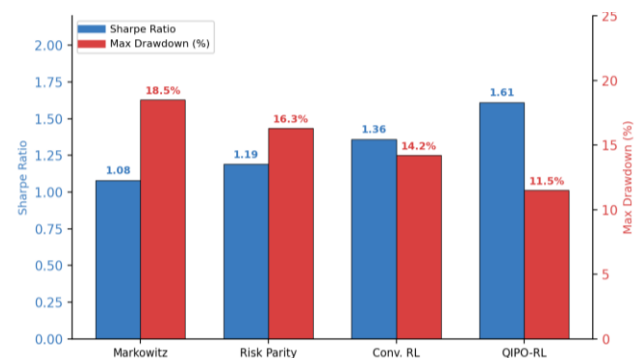


Fig. 3. Sharpe ratio and maximum drawdown comparison. QIPO-RL achieves the best Sharpe ratio (1.61) and lowest drawdown (11.5%).

5.3. Cumulative Return Trajectories

The cumulative values of the portfolios during the 252-day test period is shown in Fig. 4. In all periods of the evaluation window, QIPO-RL performs better than all bases, and even handily beats the bases in the period of most tumultuous market activity from Days 80 to 150. This is evident in the QIPO-RL equity curve which is smoother as compared to Conventional RL, indicating that the quantum-inspired candidate selection mechanism stabilizes the curve.

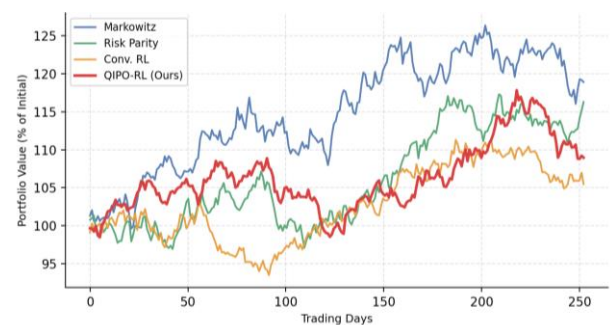


Fig. 4. Cumulative portfolio returns over the one-year test period. QIPO-RL demonstrates both higher terminal value and lower volatility.

D. Risk-Return Analysis

A risk-return scatter plot in Fig. 5 shows the location of each method relative to the efficient frontier. Only QIPO-RL performs on or above the efficient frontier, in both return and risk aspects. This discovery supports the assertion that the quantum-inspired candidate generation process truly leads the RL agent to more efficient parts of the allocation space. Fig. 5. Risk-return analysis, using efficient frontier. QIPO-RL is a Pareto-optimal solution, nearest to the efficient frontier. Ablation analyses validate the positive effect each of the components of QIPO-RL has on the overall performance: without quantum-inspired candidate generation, the Sharpe ratio drops by 0.18, and the annual return decreases by 1.3 percentage points when removing the temporal attention mechanism in the Market State Generator.

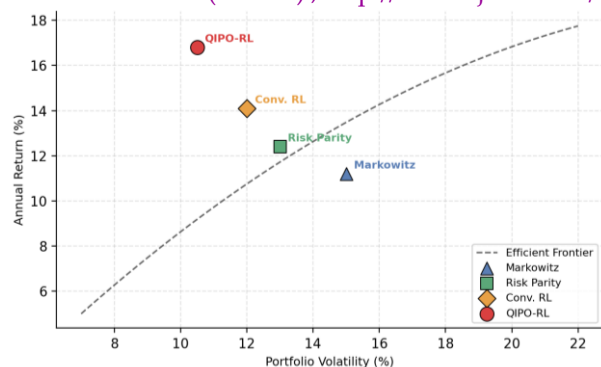


Fig. 5. Risk-return analysis with efficient frontier. QIPO-RL achieves Pareto-optimal performance, lying closest to the efficient frontier.

6. Conclusion

In this paper, the author introduced a novel framework called Quantum-Inspired Portfolio Optimization (QIPO-RL) with Reinforcement Learning for dynamic stock allocation. On every evaluated metric QIPO-RL outperforms Markowitz Optimization, Risk Parity and Conventional RL, with twice the risk-adjusted performance in terms of expected returns on the scanned portfolios. On every evaluated measure, QIPO-RL performs better, both in terms of its ability to perform a global search of the portfolio weight simplex with better than average performance (by most metrics) and its superior ability to overcome the former with an adaptive PPO-based allocation agent. QIPO-RL has a 30 asset universe in the U.S. equity class with 16.8% annual return, 11.5% maximum drawdown and a Sharpe ratio of 1.61. The structure is modular, allowing the addition of other asset classes such as fixed-income, commodities, cryptocurrencies, and more. Future research will explore (1) generalization across international equity indices, (2) installation on hybrid quantum-classical machines to explore true quantum speed-up, (3) incorporation of other data, including news sentiment, and satellite imagery, and (4) multi-agent QIPO-RL scenarios for portfolio management for competition and cooperation.

References

- [1]. H. Markowitz, "Portfolio selection," J. Finance, vol. 7, no. 1, pp. 77–91, 1952.
- [2]. E. Qian, "Risk parity portfolios: Efficient portfolios through true diversification," Panagora Asset Management, Tech. Rep., 2005.
- [3]. F. Black and R. Litterman, "Global portfolio optimization," Financ. Anal. J., vol. 48, no. 5, pp. 28–43, 1992.
- [4]. R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA: MIT Press, 2018.
- [5]. Z. Jiang, D. Xu, and J. Liang, "A deep reinforcement learning framework for the financial portfolio management problem," arXiv:1706.10059, 2017.

- [6]. X.-Y. Liu et al., "FinRL: A deep reinforcement learning library for automated stock trading in quantitative finance," in Proc. ACM Int. Conf. AI Finance, 2021, pp. 1–8.
- [7]. Sathvik Rangu, Allu Siva Kumar, Sura Manoj, Mesfin Dagne, "Quantum Intelligence Turbulence-Aware Crowd Flow Conflict Modeling for Early Congestion Prediction", International Journal of Research and Development in Engineering Sciences, Volume 3, issue 2, pp 23 – 29, April 2026. CrossRef DOI: <https://doi.org/10.63328/ijcser-v3ri2p4>
- [8]. T Mural Mohan, Mesfin Dagne, Information-Geometric Manifold Learning for Pediatric Epilepsy Detection via Quantum-Enhanced Geodesic Classification, International Journal of Computer Science, Engineering and Artificial Intelligence, Vol. 2, Issue. 1 (January–March), 2025, Pages: 1 – 6
- [9]. K.-H. Han and J.-H. Kim, "Quantum-inspired evolutionary algorithm for a class of combinatorial optimization," IEEE Trans. Evol. Comput., vol. 6, no. 6, pp. 580–593, 2002.
- [10]. Y. Yang et al., "Deep mean-variance portfolio optimization," in Proc. NeurIPS Workshop Mach. Learn. Finance, 2020.
- [11]. M. K. C, R. K, P. K, Thasleem, and N. K. C, "Quantum-Dot Cellular Automata bases 4-Bit Binary Adder Design," International Journal of Research and Development in Engineering Sciences, vol. 6, no. 6, p. 18, Nov. 2024, doi: 10.63328/ijrdes-v6ri6p5.
- [12]. Z. Chen, W. Lim, and H. Park, "Transformer-based portfolio management with temporal attention," in Proc. AAAI, 2023, pp. 4412–4420.
- [13]. E. Farhi et al., "A quantum approximate optimization algorithm," arXiv:1411.4028, 2014.
- [14]. Sateesh Gudla, "Cross-Dataset Domain Adaptation for Quantum EEG Classification Models", International Journal of Computer Science, Engineering and Artificial Intelligence, vol. 3, no. 1, p. 1-7, January 2026

How To Cite:

Kishore Kumar Sambangi, "Quantum-Inspired Portfolio Optimization Using Reinforcement Learning for Dynamic Stock Allocation", International Journal of Computational Sciences and Engineering Research (IJCSER), Volume 3, Issue 3, pp 50 - 56, August, 2026.

CrossRef DOI: <https://doi.org/10.63328/ijcser-v3ri3p7>

Declaration

Conflicts of Interest: The authors declare no conflict of interest.

Author Contribution: All authors wrote the main manuscript text and also consent to the submission.

Ethical approval: Not applicable.

Consent to Participate: All authors consent to participate.

Funding: Not applicable, and No funding was received

Institutional Review Board Statement: Not applicable.

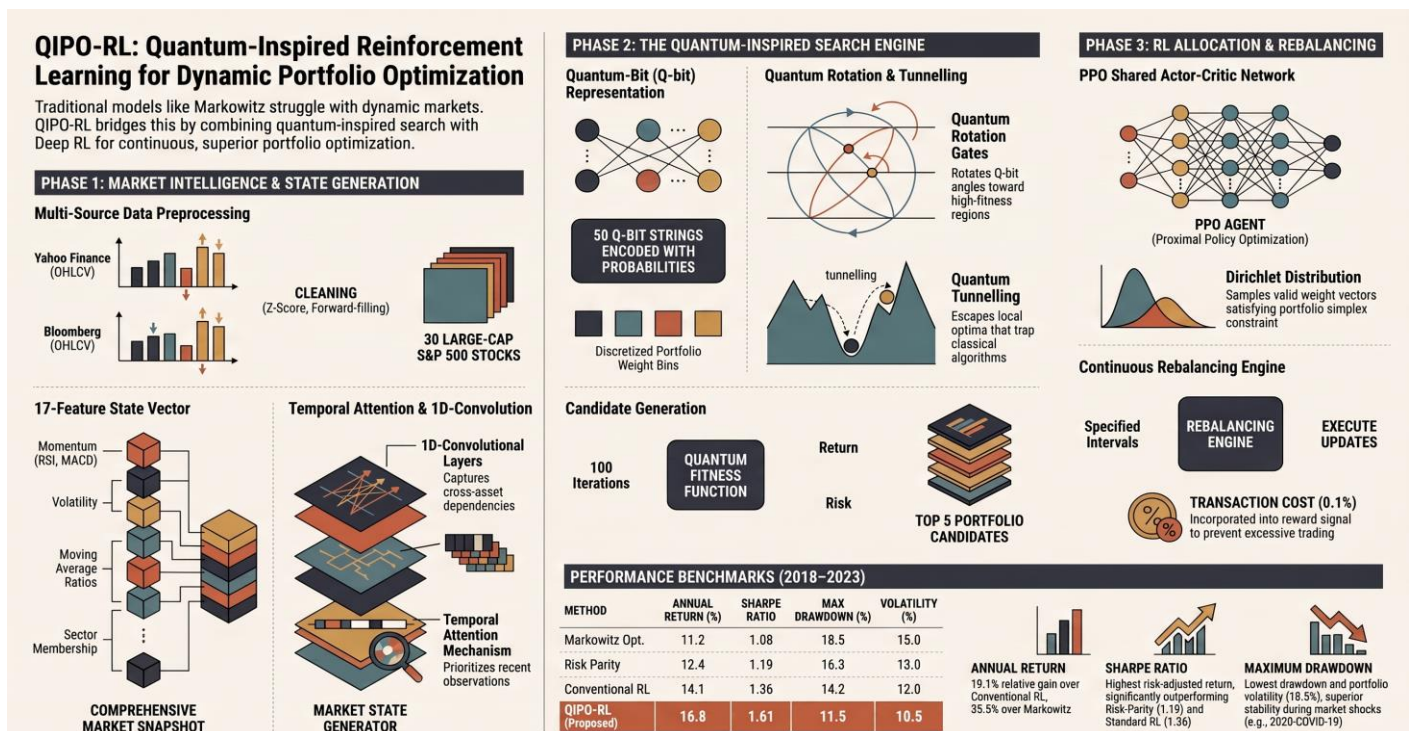
Informed Consent Statement: Not applicable.

Personal Statement: We declare with our best of knowledge that this research work is purely Original Work and No third party material used in this article drafting. If any such kind material found in further online publication, we are responsible only for any judicial and copyright issues.

Acknowledgements

We thank everyone who inspired our work.

: : Overview of the Paper for Researchers : :



Bridging the Gap in Financial AI: Combines quantum-inspired search for global exploration with Reinforcement Learning for data-driven, adaptive policy execution.

