Seed Sensitivity and Risk Efficiency in PPO-Based Portfolio Allocation: A Reproducible Diagnostic Study

hakim DJOMO GOÏMBA

hakim.djomo@aivancity.education

aivancity

Research Article

Keywords: financial reinforcement learning, portfolio allocation, PPO, seed sensitivity, reproducibility, algorithmic trading, risk diagnostics

Posted Date: May 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9824202/v1

Additional Declarations: The authors declare no competing interests.

Seed Sensitivity and Risk Efficiency in PPO-Based Portfolio Allocation: A Reproducible Diagnostic Study

Hakim Djomo Goimba
Aivancity School for Technology, Business & Society
hakim.djomo@aivancity.education

May 26, 2026

Abstract

This paper presents **RiskLens Trader**, a reproducible diagnostic study of seed sensitivity and risk efficiency in Proximal Policy Optimization (PPO)-based portfolio allocation. Rather than proposing a new trading algorithm, the study evaluates whether a standard PPO agent remains competitive under a stricter experimental protocol involving multiple random seeds, matched transaction costs, classical non-reinforcement-learning baselines, and uncertainty estimates.

The study considers long-only allocation across five large-cap US equities (AAPL, MSFT, NVDA, AMZN, and GOOGL) from 2018-01-02 to 2026-03-11, using a chronological 80/20 train-test split. PPO is trained for 30,000 timesteps under five random seeds and compared against equal-weight, buy-and-hold, momentum, and minimum-variance strategies. PPO achieves a higher mean total return than equal-weight allocation ($48.2\% \pm 51.0\%$ versus 33.6%), but its mean Sharpe ratio (0.79 ± 0.69) is lower than all four classical baselines, and its average maximum drawdown is worse than every comparator.

The best PPO seed achieves a 132.6% total return and a Sharpe ratio of 1.78, while another seed loses money out of sample. This dispersion shows that single-seed reporting can materially overstate the performance of financial reinforcement-learning systems. The main contribution is methodological: RiskLens Trader provides a compact and inspectable framework for auditing financial RL experiments through seed-level reporting, multi-baseline comparison, transaction-cost-aware evaluation, bootstrap uncertainty, drawdown analysis, and trading-behavior diagnostics. The results support adopting multi-seed and multi-baseline evaluation as a minimum reporting standard before claiming outperformance in financial reinforcement learning.

Keywords: financial reinforcement learning; portfolio allocation; PPO; seed sensitivity; reproducibility; algorithmic trading; risk diagnostics.

1 Introduction

Deep reinforcement learning (RL) has been widely explored for automated trading and portfolio allocation. In this setting, an agent observes market features, selects portfolio weights, and receives rewards based on realized returns net of trading costs. Despite the appeal of this formulation, financial RL remains difficult to evaluate: market histories are short, returns are noisy, regimes are non-stationary, and apparent outperformance can result from favorable random initialization or weak baselines.

This paper argues that the central question for small-universe financial RL is not whether a single trained agent can produce an attractive backtest, but whether the result survives basic

diagnostic scrutiny. In particular, a credible evaluation should answer three questions: (i) does performance remain stable across random seeds, (ii) does the learned policy outperform simple non-RL baselines under matched transaction costs, and (iii) are uncertainty and drawdown behavior reported rather than hidden by aggregate returns?

Portfolio construction has traditionally been formulated as a risk-return optimization problem [1]. RL re-frames allocation as sequential decision-making: an agent observes market states, adjusts portfolio weights, and receives rewards based on realized returns net of trading frictions. This has motivated applications of Deep Q-Networks [2], policy-gradient methods, and Proximal Policy Optimization (PPO) [3] to automated trading [4, 6, 7].

Prior work on deep RL has shown that random seeds and implementation details can strongly affect reported results [11]. This issue is especially severe in finance, where the signal-to-noise ratio is low, and evaluation is often performed on a single chronological split. Yet many financial RL studies still emphasize single-run backtests and comparisons against only equal-weight or buy-and-hold baselines.

This paper introduces **RiskLens Trader**, a compact diagnostic framework for evaluating PPO-based long-only portfolio allocation under a stricter reporting protocol. The objective is not to introduce a new policy-gradient method, but rather to make instability visible through seed-level reporting, classical baselines, transaction-cost-aware metrics, bootstrap uncertainty estimates, and risk-inspection visualizations.

This study should be interpreted as a diagnostic negative result rather than as a proposal for a new trading algorithm. It makes three primary contributions:

1. We provide a transaction-cost-aware diagnostic protocol for auditing financial RL allocation experiments through seed-level reporting, matched classical baselines, uncertainty estimates, drawdown analysis, and realized trading-behavior diagnostics.

2. We apply this protocol to a standard PPO allocator and show that apparent outperformance is highly sensitive to random initialization: the best seed strongly outperforms all baselines, while the multi-seed aggregate does not dominate simple classical strategies on risk-adjusted metrics.

3. We quantify how turnover and concentration mediate this instability: favorable PPO runs are associated with far higher reallocation intensity and portfolio concentration than diversified non-RL baselines.

The key empirical finding is negative but precisely quantified: in this constrained setting, PPO can produce strong individual runs, but its aggregate risk-adjusted performance is not superior to simple classical alternatives. This finding makes a concrete case for multi-seed, multi-baseline evaluation as a minimum standard in financial RL research.

2 Related Work

2.1 Portfolio Optimization

Markowitz [1] introduced mean-variance optimization, formalizing portfolio construction as a trade-off between expected return and variance. Although influential, the classical framework is sensitive to estimation error and non-stationarity. The minimum-variance portfolio, which minimizes variance without an expected-return target, is empirically more robust than the full mean-variance solution and serves as one of our baselines [12]. DeMiguel et al. [12] showed that the equal-weight

portfolio can outperform a wide range of optimized strategies out of sample, motivating its inclusion as a strong baseline.

2.2 Deep Reinforcement Learning for Trading

Deep RL has been widely studied for sequential financial decision-making. Jiang et al. [6] proposed the EIIE architecture for portfolio management. FinRL [4] provided a standardized library for deep RL in automated trading. Yang et al. [7] studied ensemble deep RL strategies and highlighted that no single method dominates across market regimes.

A critical gap in this literature is the practice of reporting favorable single-run backtests without sufficient seed-level diagnostics. Henderson et al. [11] demonstrated that RL benchmark results can be strongly affected by seed selection and implementation details in non-financial settings. RiskLens Trader provides a focused financial case study showing that the same concern arises in PPO-based portfolio allocation.

RiskLens Trader is related to FinRL [4] in scope, but differs in objective. Rather than supporting many algorithms and markets, it provides diagnostic instrumentation for a narrow and fully inspectable setting: multi-seed reporting, baseline comparison, bootstrap uncertainty, drawdown analysis, and allocation visualization.

2.3 LLMs for Financial Decision Support

Large language models have been explored for financial signal extraction and decision support [8–10]. The codebase also contains an optional LLM signal-fusion endpoint, but this component is not evaluated in the present work and is left for future ablation studies. No empirical claim in this paper depends on LLM-assisted allocation.

3 Diagnostic Evaluation Framework

RiskLens Trader is structured around two core components.

Training and evaluation layer. The computational engine is implemented in `src/evaluate.py` and `src/rl_training.py`. The RL module implements PPO training via Stable-Baselines3 [5] with multi-seed support. The evaluation module computes portfolio curves, benchmark curves, drawdowns, risk-adjusted metrics, turnover diagnostics, and bootstrap confidence intervals for reported statistics.

Serving and visualization layer. The implementation includes a FastAPI backend and a browser dashboard for inspecting NAV curves, drawdowns, allocations, seed dispersion, and baseline comparisons. These components are auxiliary to the evaluation protocol and are released with the accompanying code repository.

4 Experimental Design

4.1 Dataset and Split

The dataset `data/processed/yfinance_prices.csv` contains 2,058 daily observations for five large-cap US equities (AAPL, MSFT, NVDA, AMZN, GOOGL) from 2018-01-02 to 2026-03-11,

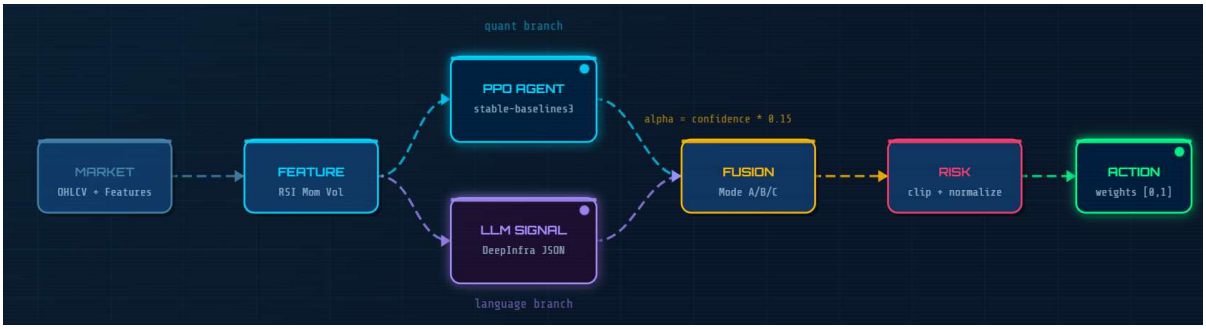

Figure 1: System architecture of RiskLens Trader. Market data are transformed into features, passed through PPO and classical baseline branches, and then evaluated with matched transaction costs, risk metrics, bootstrap uncertainty, and visualization components.

aligned on common trading dates. A chronological 80/20 split yields 1,646 training dates (2018-01-02 to 2024-07-18) and 412 test dates (2024-07-19 to 2026-03-11).

This universe is intentionally small for transparency. All five assets are US large-cap technology-oriented equities, creating sector concentration. The test window coincides with a specific market regime; results may not generalize to other sectors, asset classes, or time periods.

4.2 Feature Engineering

For each asset, the pipeline computes five daily features: one-day return, five-day return, 20-day momentum, 20-day rolling volatility, and relative gap to the 20-day moving average. All features are computed from close prices only. Feature values at time t are lagged by one day so that allocations use only information available before execution, reducing look-ahead bias.

4.3 Portfolio Environment

At time step t, the observation is the concatenation of the per-asset feature tensor, current portfolio weights, and current net asset value (NAV). For $N = 5$ assets and $F = 5$ features per asset, the observation vector has $N \cdot F + N + 1 = 31$ entries.

Actions are non-negative weight vectors clipped to $[0, 1]$ and normalized to sum to one, enforcing a long-only, fully invested portfolio.

The one-step reward is

$$r_t = w_{t-1}^{\top}\left(\frac{p_t}{p_{t-1}} - 1\right) - c \cdot \text{turnover}_t,$$

where w_{t-1} are portfolio weights, p_t asset prices, and $c = 0.001$ (10 basis points per unit of turnover).

4.4 Learning Algorithm

Training uses PPO [3] via Stable-Baselines3 [5] with an MLP policy. Configuration: learning rate 3×10^{-4}, discount factor $\gamma = 0.99$, entropy coefficient 0.01, 256 rollout steps, and batch size 64. All experiments use five random seeds (42, 123, 456, 789, and 999) and are trained for 30,000 timesteps.

4.5 Classical Baselines

We compare PPO against four classical strategies, all evaluated on the same out-of-sample dates and with the same 10 bp transaction-cost assumption applied at each rebalancing:

- **Equal-weight (EW):** 20% allocation to each asset, rebalanced daily.
- **Buy-and-hold (BH):** Equal-weight allocation at the start of the test period, never rebalanced.
- **Momentum (MOM):** At each rebalancing date, allocate proportionally to the softmax of each asset's trailing 20-day return.
- **Minimum-variance (MV):** Solve the long-only minimum-variance problem using the trailing 60-day covariance matrix, rebalanced monthly. This is a classical risk-controlled comparator.

These baselines span passive, rule-based, and optimized allocation approaches, providing a more informative comparison than a single equal-weight baseline.

4.6 Metrics

The evaluation reports total return, annualized return, Sharpe ratio, Sortino ratio, annualized volatility, maximum drawdown, and Calmar ratio on the out-of-sample test set.

We additionally report trading-behavior diagnostics. Average daily turnover is defined as

$$\text{Turnover}_t = \frac{1}{2}\sum_{i=1}^{N} |w_{t,i} - w_{t-1,i}|,$$

and annualized turnover is 252 times the daily average. Portfolio concentration is measured using the Herfindahl-Hirschman Index,

$$\text{HHI}_t = \sum_{i=1}^{N} w_{t,i}^2,$$

where larger values indicate more concentrated portfolios.

4.7 Bootstrap Confidence Intervals

To quantify uncertainty from seed selection, we report 95% bootstrap confidence intervals (CIs) for PPO metrics. For each metric m, we resample the five per-seed values with replacement for 10,000 iterations and report the 2.5th and 97.5th percentiles. Because the interval is estimated from only five seed-level observations, we interpret it as a descriptive instability diagnostic rather than a formal inferential guarantee [13].

5 Results

5.1 Multi-Baseline Comparison

Table 1 reports out-of-sample metrics for PPO (mean ± standard deviation across five seeds, with descriptive 95% bootstrap intervals) and for all four classical baselines.

Table 1: Out-of-sample evaluation metrics across PPO and four classical baselines. PPO results are reported as mean ± standard deviation across five seeds. Bracketed intervals are descriptive bootstrap intervals over five observed seed-level values and should not be interpreted as population-level confidence intervals. Bold indicates the best value per metric.

Metric	PPO (mean ± std) [95% interval]	EW	BH	MOM	MV
Total return	48.2%±51.0% [−11%, 133%]	33.6%	36.1%	41.5%	**52.3**%
Annual return	25.5%±26.4% [−6%, 69%]	19.4%	20.7%	23.5%	**27.8**%
Sharpe ratio	0.79±0.69 [0.10, 1.48]	0.83	0.81	**0.94**	0.91
Sortino ratio	0.83±0.74 [0.11, 1.56]	0.80	0.78	**0.99**	0.87
Volatility (ann.)	31.4%±3.7% [25%, 36%]	25.4%	25.4%	26.3%	**22.1**%
Max drawdown	−30.1%±1.5% [−32%, −29%]	−26.5%	−26.5%	−24.8%	**−19.3**%
Calmar ratio	0.85±0.91 [−0.25, 2.34]	0.73	0.78	0.95	**1.44**

Table 2: Per-seed out-of-sample results. Seed 456 has the highest test Sharpe ratio in this sample; seed 999 is the only loss-making run. The 144.0 percentage-point spread between the best and worst total returns illustrates the fragility of initial conditions.

Seed	Total return	Sharpe	Max drawdown	Calmar
42	10.1%	0.34	−29.4%	0.21
123	74.3%	1.29	−31.0%	1.31
456	132.6%	1.78	−29.0%	2.34
789	35.5%	0.69	−32.5%	0.63
999	−11.4%	−0.16	−28.6%	−0.25
Mean	48.2%	0.79	−30.1%	0.85
Std	51.0%	0.69	1.5%	0.91
95% interval	[−11%, 133%]	[0.10, 1.48]	[−32%, −29%]	[−0.25, 2.34]

PPO achieves a higher mean total return than equal-weight, buy-and-hold, and momentum, but it does not dominate the baseline set. Minimum-variance achieves the highest total return and annual return in this test window, while momentum achieves the highest Sharpe and Sortino ratios. PPO has the highest annualized volatility and the deepest average drawdown.

5.2 Per-Seed Results

Table 2 reports individual out-of-sample metrics for each seed.

The spread across seeds is the central empirical result. A researcher selecting only seed 456 would report a total return of 132.6% and a Sharpe ratio of 1.78, which would appear to strongly outperform all baselines. The multi-seed protocol reveals that this is a favorable initialization rather than a reliable aggregate effect.

5.3 Best-Seed Selection Bias

The five-seed experiment illustrates the magnitude of selection bias induced by reporting only the most favorable run. The best PPO seed achieves 132.6% total return and Sharpe 1.78, while the across-seed mean is 48.2% total return and Sharpe 0.79. Reporting the best seed instead of the mean, therefore, inflates total return by 84.4 percentage points and Sharpe by 0.99. The worst seed loses 11.4% out of sample. This dispersion is larger than the performance gap between the

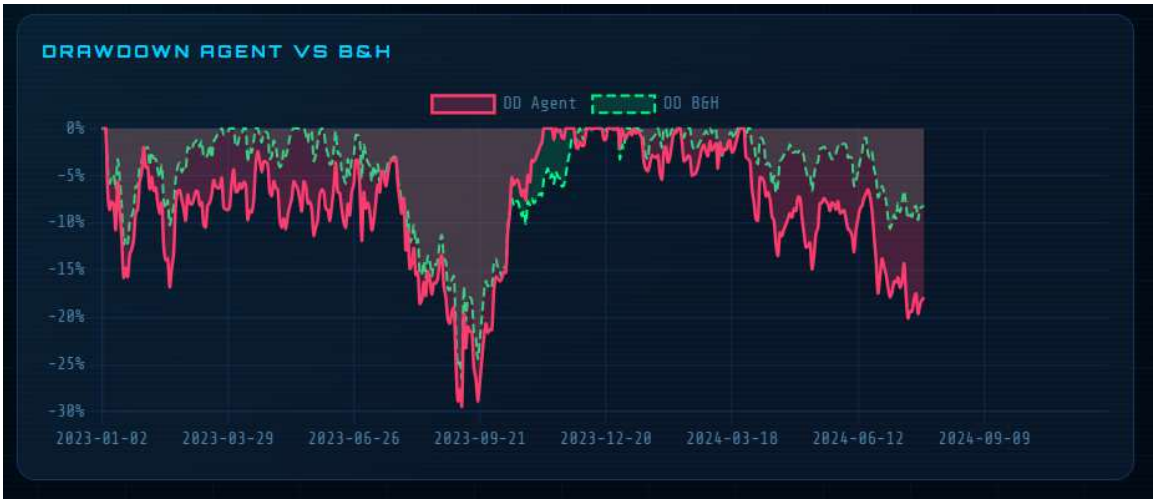

Figure 2: Out-of-sample drawdown curve for all baselines and seed 456, the highest-test-Sharpe PPO run in the reproduced five-seed sample. The minimum-variance baseline exhibits the shallowest drawdown profile. This figure is descriptive only and should not be interpreted as representative of mean PPO performance across seeds.

Table 3: Trading behavior diagnostics on the out-of-sample test set. PPO values are averaged across five seeds.

Strategy	Daily turnover	Annual turnover	Mean HHI	Max weight
PPO	0.305	76.86	0.590	0.961
EW	0.005	1.30	0.200	0.200
BH	0.000	0.00	0.203	0.280
MOM	0.005	1.35	0.201	0.244
MV	0.0003	0.08	0.200	0.202

PPO mean and any single classical baseline, indicating that seed selection can dominate algorithm-baseline comparisons in this setting.

5.4 Drawdown Analysis

Average PPO drawdown (-30.1%) is worse than every classical comparator, including passive buy-and-hold (-26.5%). Minimum-variance achieves the shallowest drawdown (-19.3%) and the highest Calmar ratio (1.44), indicating materially stronger downside control in this test window. This pattern suggests that the higher PPO returns observed in favorable seeds are not driven by uniformly better risk efficiency, but rather by more fragile loss behavior.

5.5 Trading Behavior Diagnostics

Table 3 reports realized trading-behavior diagnostics on the out-of-sample test window. PPO values are averaged across the five evaluated seeds, while the classical baselines are deterministic. These diagnostics complement return-based metrics by showing how aggressively each strategy reallocates capital and how concentrated the resulting portfolios become over time.

The diagnostics show that PPO behaves very differently from the classical baselines. Its mean daily turnover is 0.305, corresponding to an annualized turnover of 76.86, which is far above all

deterministic comparators. PPO is also substantially more concentrated: its mean HHI is 0.590, and its mean maximum position weight is 0.961, indicating that the learned policy often places very large bets on a single asset or a very small subset of assets.

By contrast, the classical baselines remain close to diversified allocations. Equal-weight is mechanically stable, buy-and-hold drifts only moderately from its initial diversification, momentum remains only slightly more active than equal-weight, and minimum-variance portfolios are both low-turnover and highly diversified. These diagnostics help explain why favorable PPO runs can still coincide with weaker drawdown control: part of the achieved upside is associated with aggressive reallocation and high concentration rather than with stable risk-efficient diversification.

5.6 Interpretation

PPO does not dominate classical baselines. The five-seed bootstrap interval for the PPO Sharpe ratio is wide, $[0.10, 1.48]$, indicating substantial uncertainty under seed resampling. PPO's mean Sharpe is lower than equal-weight, buy-and-hold, momentum, and minimum-variance in this test window. Its mean drawdown is also worse than all four classical strategies.

Seed dispersion remains the dominant diagnostic signal. The difference between the best and worst PPO seeds is larger than the difference between the PPO mean return and any baseline return. A single favorable seed, such as seed 456, would support a strong outperformance narrative, but the multi-seed view shows that this conclusion is not robust to initialization.

Turnover and concentration clarify the mechanism. The trading-behavior diagnostics show that PPO's performance is associated with a much more aggressive allocation style than any classical comparator. The policy turns over capital far more frequently and concentrates risk much more heavily, often approaching single-name exposure. This makes the PPO policy economically different from the diversified baselines and helps explain why higher returns in favorable seeds do not translate into consistently superior risk-adjusted performance.

Classical baselines remain competitive. Momentum achieves the highest Sharpe and Sortino ratios without training cost or random initialization, while minimum-variance delivers the best drawdown control and the highest Calmar ratio. These results reinforce the importance of including classical non-RL comparators when evaluating financial RL systems.

6 Threats to Validity

External validity. The study uses five large-cap US technology-oriented equities. This narrow universe allows full traceability but limits generalization to other sectors, asset classes, liquidity regimes, and market conditions. The results should therefore be interpreted as a diagnostic case study rather than as a universal claim about PPO in finance.

Temporal validity. The evaluation uses a single chronological train-test split. Although this avoids random temporal leakage, it does not establish robustness across market regimes. Walk-forward evaluation over multiple test windows is necessary before drawing stronger conclusions.

Statistical validity. Only five PPO seeds are evaluated. Bootstrap intervals over five seed-level observations are useful for describing instability, but should not be treated as strong inferential guarantees. Future work should increase the number of seeds and report uncertainty across both seeds and time windows.

Implementation validity. The PPO configuration follows a fixed Stable-Baselines3 setup and is not extensively tuned. A different hyperparameter search could improve PPO performance, but such tuning would itself require careful validation across seeds and baselines to avoid selection bias.

Financial validity. Backtested performance does not imply deployable trading performance. The study does not model liquidity constraints, slippage beyond fixed transaction costs, taxes, borrowing constraints, market impact, or survivorship bias. The framework is intended for research diagnostics and not for investment advice.

7 Broader Impact

This work examines evaluation methodology for financial reinforcement learning, aiming to mitigate overclaiming in algorithmic trading research by increasing transparency regarding seed sensitivity, drawdowns, and baseline competitiveness. Nevertheless, financial machine learning systems may be misapplied if users interpret backtests as indicative of future profitability. Interactive dashboards could further contribute to overconfidence by presenting unstable strategies as operationally robust. Accordingly, RiskLens Trader is intended solely as a research and auditing tool rather than an investment system. Any practical deployment would necessitate extensive validation, compliance review, risk management, and human oversight.

8 Data, Code, and Materials Availability

The experiments use publicly available daily equity price data obtained through the Yahoo Finance data interface. The accompanying repository contains the application code, the FastAPI backend, the browser dashboard, the PPO training and evaluation modules, notebooks, tests, and reproduction instructions. In particular, the repository includes scripts and notebooks for regenerating the processed dataset, reproducing the PPO multi-seed evaluation, computing the classical baselines, and generating the reported tables and figures.

The public code repository is available at:

`https://github.com/hakimoney/risklens-trader`

The repository is organized around three components: (i) a FastAPI backend for evaluation endpoints and dashboard serving, (ii) a frontend dashboard for inspecting performance, comparisons, multi-seed behavior, allocation, risk, insights, and optional LLM outputs, and (iii) a computation layer implementing portfolio metrics, PPO training with Stable-Baselines3, multi-symbol training, and multi-seed evaluation.

If redistribution of downloaded market data is restricted by the data provider's terms of use, the repository should provide data-generation scripts rather than redistributing raw market data directly. The LLM-related endpoint is optional and is not required to reproduce the empirical claims in this paper.

9 Compute Resources

All experiments were run using Python 3.11. Each PPO seed was trained for 30,000 timesteps, and the full multi-seed evaluation used the five random seeds reported in Section 4. The repository provides the required Python dependencies in `requirements.txt` and includes tests that can be executed with `pytest -q`. The dashboard can be launched locally with `uvicorn app.server:app -reload`.

The experiments were designed to be reproducible on a standard personal workstation. For a public archival version, the author should provide exact machine details when available, including operating system, CPU model, GPU model (if used), RAM, package versions, approximate runtime per seed, and total runtime for the full multi-seed experiment.

10 Disclaimer

This paper is provided for research and educational purposes only. It does not constitute investment advice, financial advice, or a recommendation to buy, sell, or hold any financial asset. Backtested performance does not imply future performance.

11 Conclusion

This paper presented RiskLens Trader, a reproducible diagnostic study of seed sensitivity and risk efficiency in PPO-based portfolio allocation. The objective is not to propose a new trading algorithm, but to evaluate whether a standard PPO agent remains competitive under a stricter reporting protocol involving multiple seeds, matched transaction costs, classical baselines, uncertainty estimates, and drawdown analysis.

In the studied five-asset setting, PPO produces highly variable outcomes. The best seed suggests strong outperformance, while another seed loses money out of sample. In aggregate, PPO achieves a higher mean total return than equal-weight but a lower mean Sharpe ratio than equal-weight, buy-and-hold, momentum, and minimum-variance baselines. Its average maximum drawdown is also worse than every classical comparator. The realized trading diagnostics also show that PPO relies on much higher turnover and much stronger concentration than the deterministic baselines. Collectively, these results demonstrate that single-seed financial RL reporting can materially overstate performance.

The study is intentionally limited in scope and should not be interpreted as a universal statement about PPO or financial RL. Its purpose is to demonstrate, in a fully inspectable setting, how an apparently strong single-run result can be reversed by standard robustness diagnostics. The broader implication is methodological: financial RL research should report seed-level dispersion, multiple classical baselines, transaction-cost stress tests, uncertainty estimates, and realized trading-behavior diagnostics before claiming outperformance.

References

[1] Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77–91.

[2] Mnih, V., Kavukcuoglu, K., Silver, D., et al. (2015). Human-level control through deep reinforcement learning. *Nature*, 518(7540), 529–533.

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

[4] Liu, X.-Y., Yang, H., Chen, Q., et al. (2020). FinRL: A deep reinforcement learning library for automated stock trading in quantitative finance. *arXiv preprint arXiv:2011.09607*.

[5] Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., & Dormann, N. (2021). Stable-Baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, 22(268), 1–8.

[6] Jiang, Z., Xu, D., & Liang, J. (2017). A deep reinforcement learning framework for the financial portfolio management problem. *arXiv preprint arXiv:1706.10059*.

[7] Yang, H., Liu, X.-Y., Zhong, S., Walid, A. (2020). Deep reinforcement learning for automated stock trading: An ensemble strategy. In *Proceedings of the First ACM International Conference on AI in Finance*.

[8] Wu, S., Irsoy, O., Lu, S., et al. (2023). BloombergGPT: A large language model for finance. *arXiv preprint arXiv:2303.17564*.

[9] López-Lira, A., & Tang, Y. (2023). Can ChatGPT forecast stock price movements? Return predictability and large language models. *arXiv preprint arXiv:2304.07619*.

[10] Benhenda, M. (2025). FinRL-DeepSeek: LLM-Infused Risk-Sensitive Reinforcement Learning for Trading Agents. *arXiv preprint arXiv:2502.07393*.

[11] Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., & Meger, D. (2018). Deep reinforcement learning that matters. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1).

[12] DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies*, 22(5), 1915–1953.

[13] Efron, B., & Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. Chapman & Hall/CRC.

A Reproducibility Details

The recommended local environment uses Python 3.11. The accompanying repository contains scripts and notebooks to reproduce the reported PPO multi-seed evaluation, classical baselines, tables, figures, and tests.

```
git clone https://github.com/hakimoney/risklens-trader.git
cd risklens-trader
python -m venv .venv
pip install -r requirements.txt
pytest -q
```

On Windows PowerShell, the virtual environment can be activated with:

```
.\.venv\Scripts\Activate.ps1
```

The dashboard can be launched locally with:

```
uvicorn app.server:app --reload
```

The application then runs locally at `http://127.0.0.1:8000`. Optional LLM functionality requires a DeepInfra API key, but this component is not required to reproduce the paper's reported PPO and baseline results.

B Transparency Checklist

This preprint is accompanied by a transparency checklist instead of the NeurIPS-specific checklist. Before submission to OSF or SSRN, the author should verify that the public record includes: author identity and affiliation, data-source description, repository link, license information, compute-resource details, reproduction commands, optional LLM configuration notes, and the financial disclaimer.