OPAL: An Ontology-Based Framework with LSTM and PPO for Return Forecasting and Portfolio Optimization

Igor Felipe Carboni Battazza[1,2] **, Cleyton Mário de Oliveira Rodrigues**[1] **,**
João Fausto Lorenzato de Oliveira[1]

[1]Polytechnic School of Pernambuco, University of Pernambuco (POLI/UPE), Recife, Brazil
R. Benfica, 455 – Madalena – Recife – PE – 50720-001 – Brazil

[2]FITec – Technological Innovations, Recife, Brazil

{igor.fcbattazza@upe.br, cleyton.rodrigues, fausto.lorenzato}@upe.br

Abstract. *Traditional portfolio optimization struggles with market volatility and nonlinear dynamics. We propose OPAL, a hybrid framework integrating semantic reasoning and machine learning for financial decision-making. OPAL combines ontology-based asset selection, Markowitz diversification, LSTM return forecasting, and PPO-based allocation. Using S&P 500 data (2015–2023), OPAL achieves a 118.05% cumulative return, 2.58 Sharpe ratio, and 4.99 Sortino ratio, surpassing baselines. Its modular design ensures transparency and adaptability.*

1. Introduction

The increasing complexity and volatility of financial markets, driven by macroeconomic uncertainty and high-frequency trading, has challenged traditional approaches to portfolio management. Classical methods such as mean-variance optimization [Markowitz 1952], despite their theoretical elegance, rely on restrictive assumptions like linearity, stationarity, and normally distributed returns. These limitations hinder their ability to adapt to real-world market dynamics, which are inherently nonlinear, temporally dependent, and influenced by a wide array of structured and unstructured information. In response, the financial machine learning community has increasingly embraced data-driven methods capable of capturing hidden temporal patterns and modeling sequential decision-making under uncertainty.

While recent advances in deep learning and reinforcement learning have significantly improved the modeling of return dynamics and policy optimization, these models often lack explainability and rely solely on quantitative signals extracted from historical prices. On the other hand, approaches based on semantic reasoning, such as ontologies, offer a structured way to incorporate domain knowledge, formalize expert rules, and promote model interpretability. However, these symbolic systems are typically used in isolation from learning algorithms, which limits their responsiveness to dynamic market contexts. This disconnect highlights a critical gap: although hybrid models have emerged in the literature, few solutions effectively integrate symbolic reasoning, deep forecasting, and adaptive policy learning into a coherent and operational pipeline for financial decision support.

To address this gap, we propose **OPAL** (Ontology-based Portfolio Allocation with LSTM and PPO), a hybrid and modular framework that unifies semantic asset filtering, deep return forecasting, and reinforcement learning for dynamic portfolio optimization. OPAL is composed of four interconnected stages: (i) asset selection based on financial ontologies with dynamically computed thresholds across multiple economic indicators; (ii) portfolio pre-selection using a Markowitz filter to ensure diversification and control risk exposure; (iii) return forecasting via Long Short-Term Memory (LSTM) networks, which model temporal dependencies in price data; and (iv) portfolio allocation using Proximal Policy Optimization (PPO), a robust deep reinforcement learning algorithm capable of learning dynamic policies under noisy environments. This design allows the agent to learn investment strategies that are not only responsive to recent market conditions but also guided by fundamental financial reasoning.

The core novelty of OPAL lies in its ability to reconcile symbolic interpretability with empirical learning efficiency. Unlike existing models that treat asset selection, return prediction, and allocation as separate tasks, OPAL introduces a tightly integrated architecture in which ontological reasoning enriches the information set available to the forecasting model and reinforcement agent. This integration leads to enhanced policy learning, greater transparency in asset selection, and improved performance across multiple evaluation metrics.

The main contributions of this work are threefold: (i) we introduce a novel framework that combines ontological reasoning, LSTM-based forecasting, and reinforcement learning for financial portfolio management; (ii) we demonstrate the effectiveness of this hybrid architecture through extensive experiments on real-world S&P 500 data, showing superior risk-adjusted performance compared to traditional and machine learning-based baselines; and (iii) we provide a modular and extensible design that supports the incorporation of alternative predictive models, risk metrics, and optimization objectives.

The remainder of this paper is organized as follows: Section 2 discusses related work. Section 3 details the OPAL framework. Section 4 presents the experimental setup and results. Finally, Section 5 provides conclusions and future research directions.

2. Related Work

Portfolio optimization is a long-standing problem in finance, traditionally addressed by the seminal work of Harry Markowitz in [Markowitz 1952], who introduced the mean-variance theory. Despite its elegance, this method relies on strong assumptions such as stationarity and Gaussianity of returns, which are often violated in real markets.

2.1. Deep Learning for Financial Forecasting

Deep learning models have gained popularity in forecasting financial time series due to their ability to capture nonlinear temporal dependencies. Long Short-Term Memory (LSTM) networks [Hochreiter and Schmidhuber 1997] have been widely adopted in this context, demonstrating superior performance over autoregressive models like ARIMA in various studies [Fischer and Krauss 2018, Siami-Namini et al. 2018]. Recent works have also explored hybrid deep learning architectures, such as attention-based LSTMs [Qin et al. 2017], and transformer-based models [Wu 2020], to further enhance predictive accuracy.

2.2. Reinforcement Learning for Portfolio Management

Reinforcement Learning (RL) has shown great promise in modeling sequential decision-making in portfolio allocation. Early works applied Q-learning and SARSA to trading environments [Moody and Saffell 1998, Neuneier 1998]. With the advent of Deep Reinforcement Learning (DRL), algorithms such as DQN [Mnih 2015], DDPG [Lillicrap 2015], A3C [Mnih 2016], and PPO [Schulman 2017] have been successfully employed in finance [Jiang et al. 2017, Yu et al. 2019]. PPO stands out for its robustness and sample efficiency, making it suitable for high-dimensional, noisy financial environments.

2.3. Ontologies and Semantic Filtering in Finance

Semantic filtering through ontologies allows the formalization of domain knowledge for asset selection. Bhuyan and Sastry [Bhuyan and Sastry 2023] proposed an ontology-based framework that encodes financial ratios and decision rules to classify stocks into performance categories. Ontological reasoning enables explainability and adaptability, making it a strong complement to data-driven models.

Other works have leveraged ontologies for portfolio construction [Reis et al. 2017], credit scoring [Albadra and Iqbal 2019], and financial event analysis [Buitelaar et al. 2005], showing that structured semantic knowledge can improve interpretability and screening effectiveness.

2.4. Integrated Hybrid Architectures

[Liu et al. 2022] introduced an integrated framework combining BERT-based sentiment analysis, Non-Stationary Markov Chains (NMC), LSTM, and Deep Q-Learning (DQN) for quantitative trading. Similarly, [Ye et al. 2020] proposed a hierarchical reinforcement learning framework that incorporates expert signals and neural networks for trading strategies.

While these works highlight the power of hybrid systems in financial decision-making, they typically combine deep learning models and reinforcement learning with limited or no integration of symbolic domain knowledge. For example, the framework in [Liu et al. 2022] proposed a multimodal architecture combining sentiment analysis with Non-Stationary Markov Chains and Deep Q-Learning, yet their framework lacks semantic asset screening grounded in economic indicators. Likewise, the strategy in [Ye et al. 2020] introduced a hierarchical reinforcement learning strategy that incorporates expert-derived signals, but without a formal ontological structure to encode financial reasoning. In contrast, OPAL introduces a fully integrated pipeline that bridges symbolic reasoning and empirical learning: it employs a financial ontology to guide asset selection, uses LSTM-based forecasting to capture temporal dependencies, and trains a PPO agent for dynamic reallocation based on structured and predictive inputs. This integration not only enhances transparency and interpretability but also leads to improved risk-adjusted performance.

To further illustrate the contributions of OPAL relative to these works, Table 1 provides a checklist summary, highlighting the use of key techniques across related studies.

Tabela 1. Checklist of techniques employed in related works and the proposed framework

Work	Ontology	Markowitz	LSTM	PPO (RL)
Markowitz (1952)	✗	✓	✗	✗
Hochreiter & Schmidhuber (1997)	✗	✗	✓	✗
Fischer & Krauss (2018)	✗	✗	✓	✗
Mnih et al. (2015)	✗	✗	✗	✓
Schulman et al. (2017)	✗	✗	✗	✓
Bhuyan & Sastry (2023)	✓	✗	✗	✗
Liu et al. (2022)	✗	✗	✓	✓
Siami-Namini et al. (2018)	✗	✗	✓	✗
Qin et al. (2017)	✗	✗	✓	✗
Wu et al. (2020)	✗	✗	✗	✗
Moody & Saffell (1998)	✗	✗	✗	✓
Neuneier (1998)	✗	✗	✗	✓
Lillicrap et al. (2015)	✗	✗	✗	✓
Mnih et al. (2016)	✗	✗	✗	✓
Jiang et al. (2017)	✗	✗	✓	✓
Yu et al. (2019)	✗	✗	✓	✓
Reis et al. (2017)	✓	✓	✗	✗
Ye et al. (2020)	✗	✗	✓	✓
Battazza, et al. (2025)	✓	✓	✓	✓

3. Proposed Framework and Methodology

3.1. Proposed Framework: OPAL

The OPAL framework integrates semantic reasoning, deep forecasting, and reinforcement learning for portfolio optimization. It comprises four sequential components: (i) ontology-based asset selection, (ii) Markowitz filtering, (iii) LSTM-based return forecasting, and (iv) PPO-based portfolio optimization, as detailed below. Figure 1 illustrates the high-level architecture of the OPAL framework.

3.1.1. Ontology-Based Asset Selection

The first stage employs a domain-specific financial ontology to filter out underperforming or financially unstable assets. This ontology encodes rules based on key indicators such as Current Ratio, Debt-to-Equity, Return on Equity, Revenue Growth, and others, following the logic proposed by [Bhuyan and Sastry 2023]. Unlike static knowledge bases, the ontology in OPAL is dynamically updated at each rebalance point using sector-specific medians and variances derived from the latest available data. This allows the asset screening process to adapt to evolving market contexts while maintaining semantic interpretability. Assets are ranked based on rule compliance weighted by indicator variability, and only the top-N candidates advance to the next stage.

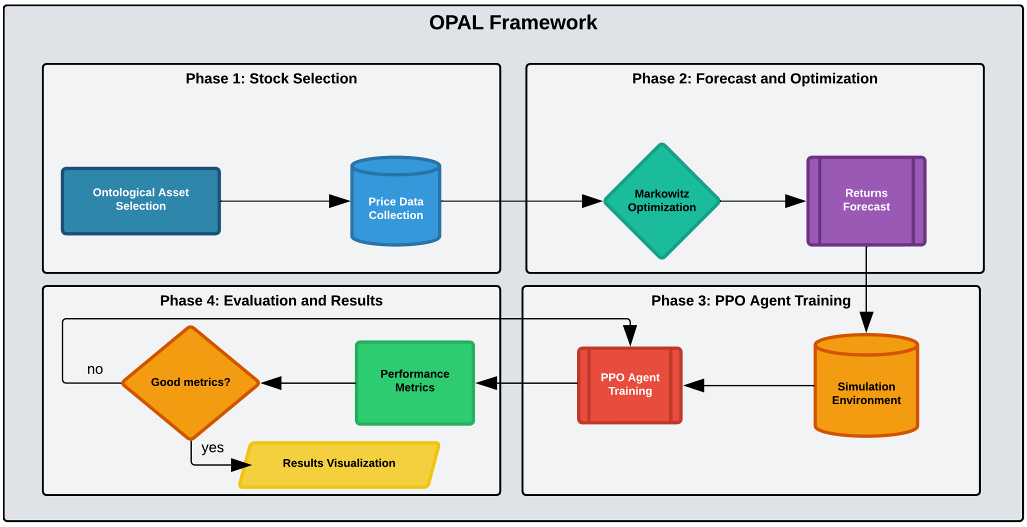

Figura 1. Overview of the OPAL framework: each module contributes to a unified learning and decision pipeline.

3.1.2. Markowitz Filtering

To ensure proper diversification and risk-aware selection, the assets retained from the ontological filter are subjected to a mean-variance optimization layer based on the classical Markowitz model [Markowitz 1952]. Expected returns and the covariance matrix are computed using historical price data, and the optimal allocation vector is obtained by solving a constrained quadratic programming problem with non-negative weights summing to one. The top-weighted assets are selected for forecasting and reinforcement learning.

3.1.3. Return Forecasting with LSTM

For each selected asset, a sliding window of historical prices is constructed, and a Long Short-Term Memory (LSTM) model [Hochreiter and Schmidhuber 1997] is trained to forecast the expected returns μ_t at each decision point t. The model is trained using normalized inputs and dropout regularization to mitigate overfitting. The output of this component is a vector of predicted returns $\mu_t \in R^n$, where n is the number of selected assets. These forecasts are passed to the next module as part of the environment's state representation.

3.1.4. Portfolio Optimization with PPO

In the final stage, OPAL employs a Proximal Policy Optimization (PPO) agent [Schulman 2017] to learn dynamic allocation strategies. The agent observes a state vector s_t composed of a concatenation of the most recent asset returns and the forecasted return vector μ_t, and outputs a continuous action vector a_t representing portfolio weights:

$$a_t = \pi_\theta(s_t) = \pi_\theta([\mathbf{r}_{t-\tau:t-1}, \mu_t]), \tag{1}$$

where π_θ denotes the policy parameterized by θ, $\mathbf{r}_{t-\tau:t-1}$ denotes the return history window of length τ, and μ_t is the predicted return vector from the LSTM. The agent is trained to maximize the following reward at each time step:

$$R_t = r_t - \alpha \cdot \mathrm{DD}_t - \beta \cdot \mathrm{TO}_t, \tag{2}$$

where r_t is the portfolio return, DD_t is the relative drawdown, and TO_t is the turnover. The coefficients α and β control the penalization of excessive risk and transaction volume, respectively. The PPO agent is implemented using the `stable-baselines3` library with entropy regularization and a clipped objective to ensure policy stability.

3.2. Research Methodology

This study adopts a design science approach, in which the OPAL framework is proposed as a computational artifact and evaluated through simulation-based experiments. Historical daily adjusted closing prices of S&P 500 constituents were collected from Yahoo Finance, covering the period from January 2015 to December 2023. A subset of the 100 most liquid stocks with complete fundamental and price data was selected to ensure realistic market conditions.

At each rebalance cycle, the framework executes the full pipeline of dynamic selection, forecasting, and allocation. An expanding window approach is used, retraining the model with accumulating historical data before each rebalance. Forecasts are made over a 20-business-day horizon. OPAL's performance is compared to traditional and machine learning-based baselines, including buy-and-hold, equal-weighted portfolios, ARIMA, LSTM with attention, and transformer-based models.

Performance is assessed using cumulative return, Sharpe ratio, Sortino ratio, maximum drawdown, and annualized volatility. Bootstrapped confidence intervals (1,000 replications) are computed for the Sharpe ratio to assess statistical robustness.

All models were implemented in Python 3.10+ using `stable-baselines3` (PPO) and `TensorFlow/Keras` (LSTM), and trained under identical hardware conditions to ensure fair comparison. Code and datasets are available upon request.

4. Experiments and Results

This section presents the experimental setup, predictive model evaluation, and portfolio performance analysis of the OPAL framework.

4.1. Experimental Setup

At each rebalance cycle, OPAL executes its full pipeline, including ontology-based selection, Markowitz filtering, return forecasting with LSTM, and portfolio optimization via PPO. The reinforcement learning agent is trained for 50,000 timesteps per cycle using the PPO implementation from `stable-baselines3`. The state vector concatenates historical returns and the LSTM-predicted return vector μ_t, allowing the policy $\pi_\theta(s_t)$ to produce allocation decisions.

The reward function is defined as:

$$R_t = r_t - \alpha \cdot \mathrm{DD}_t - \beta \cdot \mathrm{TO}_t, \tag{3}$$

where r_t is the portfolio return, DD_t is the drawdown, and TO_t is the turnover. Coefficients were set to $\alpha = 0.2$ and $\beta = 0.05$ to balance return, risk, and transaction costs.

The LSTM model consists of two layers with 64 hidden units (ReLU activation), dropout rate of 0.2, and a final dense output layer. It is trained for 10 epochs with early stopping (patience = 5) and mean squared error loss using the Adam optimizer. Assets that become unavailable during the simulation are replaced by the next top-ranked candidate based on updated financial data.

4.2. Forecasting Model Comparison

To evaluate forecasting performance, we compare OPAL's LSTM with a variety of baseline models under identical conditions. Table 2 presents the results in terms of cumulative return, Sharpe ratio, Sortino ratio, maximum drawdown (MDD), and annualized volatility.

The Sharpe ratio is a widely used metric for assessing the risk-adjusted return of an investment strategy. It is calculated as the average excess return per unit of volatility, typically annualized. A higher Sharpe ratio indicates that the strategy delivers greater returns for each unit of risk taken. In general, values above 1.0 are considered acceptable, above 2.0 are viewed as good, and above 3.0 are regarded as excellent in financial contexts.

Tabela 2. Performance metrics of return forecasting models (2015–2023)

Model	Return (%)	Sharpe	Sortino	MDD (%)	Volatility (%)
LSTM	**118.05**	2.31	**4.99**	-18.29	43.44
LSTM + Attention	60.46	**2.58**	4.20	-11.67	22.44
ARIMA	60.90	2.20	4.08	-9.13	26.92
Linear Regression	64.59	2.10	4.08	-11.28	29.81
XGBoost	65.09	2.06	3.95	-11.58	30.71
Transformer	51.15	1.87	3.70	-11.94	27.22
Momentum	48.27	1.82	3.78	-13.76	26.47
Random Forest	45.67	1.81	3.62	-13.38	25.37
Ridge Regression	41.72	1.78	3.52	-13.26	23.48
Mean	40.12	1.71	3.43	-13.10	22.97
EWMA	37.94	1.66	3.35	-13.42	22.30
GARCH	34.55	1.60	3.17	-14.80	21.54
Buy & Hold	7.99	0.47	0.79	-18.94	17.66

4.3. Discussion

The LSTM model achieves the highest cumulative return (118.05%) among all evaluated models, indicating strong ability to capture temporal price dependencies. However, it also exhibits the highest volatility and drawdown. The LSTM+Attention variant offers a better trade-off, with a Sharpe ratio of 2.58 and lower volatility. Transformer-based models show potential, but underperform LSTM in this configuration. Traditional baselines (GARCH, EWMA, ARIMA) deliver stable but less competitive results. Overall, the experiments validate the use of LSTM-based models as effective forecasting engines for guiding the reinforcement learning agent in OPAL.

5. Conclusion

This work introduced **OPAL**, a modular framework for financial portfolio optimization that integrates semantic reasoning, statistical filtering, deep forecasting, and reinforcement learning. OPAL operates in four coordinated stages: (i) semantic asset selection using dynamically updated ontological rules, (ii) diversification-aware filtering through Markowitz optimization, (iii) return forecasting with LSTM models, and (iv) portfolio allocation via PPO guided by a custom risk-aware reward function. By combining symbolic and data-driven techniques, OPAL offers an interpretable and adaptive pipeline aligned with real-world market dynamics.

Experimental results on S&P 500 data from 2015 to 2023 show that OPAL outperforms traditional heuristics and machine learning baselines across key metrics such as cumulative return, Sharpe ratio, and Sortino ratio. The integration of ontological filters with time series forecasting enhances both transparency and performance in asset allocation. Compared to prior multimodal architectures, such as the NMC-BERT-LSTM-DQN-X model [Liu et al. 2022] and hierarchical RL frameworks [Ye et al. 2020], OPAL distinguishes itself by embedding a formal ontology into the selection layer, enabling domain knowledge to influence downstream decisions.

Despite its contributions, this study has limitations. First, the evaluation is restricted to U.S.-based equities from the S&P 500, limiting generalizability to other markets or asset classes. The framework also relies on deep learning models (LSTM, PPO) that require significant computational resources, which may hinder application in environments with limited data or infrastructure. In addition, although attention-based models such as Transformers have shown potential in capturing long-range dependencies, they were not fully integrated into OPAL due to computational constraints and training instability in initial trials. Moreover, OPAL currently optimizes a single objective; incorporating multi-objective formulations (e.g., balancing risk and transaction cost) and advanced optimization techniques remains an open direction. A more rigorous sensitivity analysis, including statistical tests and experiments with varying LSTM horizons and rebalance frequencies, is also necessary to deepen the empirical validation.

Future work includes extending OPAL into a multi-objective reinforcement learning setting, incorporating criteria such as transaction cost minimization and risk-adjusted reward optimization using algorithms like A2C or PPO-Clip. We also intend to explore Transformer-based models to better capture long-term dependencies, and to integrate Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR) into both the reward function and the filtering layers for more robust risk control.

In summary, OPAL demonstrates that the fusion of structured knowledge and deep learning can produce more transparent, adaptive, and effective portfolio strategies. Its modular design opens promising avenues for future enhancement and application.

Referências

Albadra, M. and Iqbal, R. (2019). An ontology-based model for credit scoring in peer-to-peer lending. *Procedia Computer Science*, 164:520–527.

Bhuyan, B. P. and Sastry, H. (2023). Stock selection using ontological financial analysis. *International Journal of Information Technology and Management*.

Buitelaar, P., Eigner, T., and Declerck, T. (2005). Ontology-based information extraction with soba. In *LREC*.

Fischer, T. and Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2):654–669.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. In *Neural Computation*, volume 9, pages 1735–1780.

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

Lillicrap, T. e. a. (2015). Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*.

Liu, Y., Li, Y., and Sun, Y. (2022). Forecasting the market with machine learning algorithms: An application of nmc-bert-lstm-dqn-x algorithm in quantitative trading. *Mathematics*, 10(3):347.

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

Mnih, V. e. a. (2015). Human-level control through deep reinforcement learning. *Nature*, 518:529–533.

Mnih, V. e. a. (2016). Asynchronous methods for deep reinforcement learning. *ICML*.

Moody, J. and Saffell, M. (1998). Reinforcement learning for trading. *Advances in Neural Information Processing Systems*, 11:917–923.

Neuneier, R. (1998). Enhancing q-learning for optimal asset allocation. In *NIPS*.

Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G., and Cottrell, G. (2017). A dual-stage attention-based recurrent neural network for time series prediction. In *IJCAI*.

Reis, J., Silva, D., and Martinho, D. (2017). An ontology-based approach to portfolio construction. *Expert Systems with Applications*, 72:391–401.

Schulman, J. e. a. (2017). Proximal policy optimization algorithms. In *arXiv preprint arXiv:1707.06347*.

Siami-Namini, S., Tavakoli, N., and Siami Namin, A. (2018). A comparison of arima and lstm in forecasting time series. *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 1394–1401.

Wu, H. e. a. (2020). Deep transformer models for time series forecasting: A survey. *arXiv preprint arXiv:2009.08317*.

Ye, W., Xu, Y., and Lin, J. (2020). Financial portfolio allocation with hierarchical reinforcement learning. *IEEE Access*, 8:123208–123218.

Yu, L., Zhang, H., Feng, Y., and Xu, X. (2019). Deep reinforcement learning for portfolio management. *Complexity*, 2019.