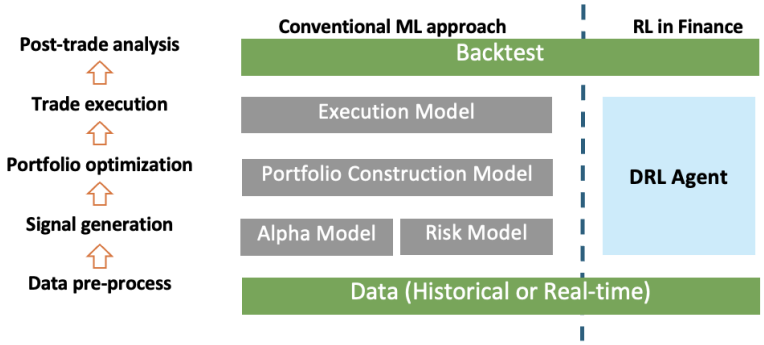
Financial Reinforcement Learning

Algorithmic trading plays an important role in financial markets, accounting for over 60% of the total trading volume in the U.S. equity market. It automates the decision-making process, which requires dynamically deciding when to trade, what assets to trade, at what price, and in what quantity. These decisions must be made in real-time within highly volatile and uncertain market environments.

Conventional machine learning (ML) approaches to algorithmic trading typically rely on a multi-stage pipeline. It involves multiple models, such as alpha models to generate signals, risk models for risk modeling, portfolio construction models to allocate capital, and execution models to perform trading strategies. These models often need to be optimized independently and then manually coordinated, which may ignore the nature of sequential trading decisions and have difficulty adapting to changing market conditions. **Reinforcement Learning (RL),** specifically **Deep RL (DRL),** flips the script by training a single model to learn optimal trading strategies through trial and error—just like a human trader, but at machine learning speed and scale. It simplifies the workflow while adapting to dynamic market conditions.

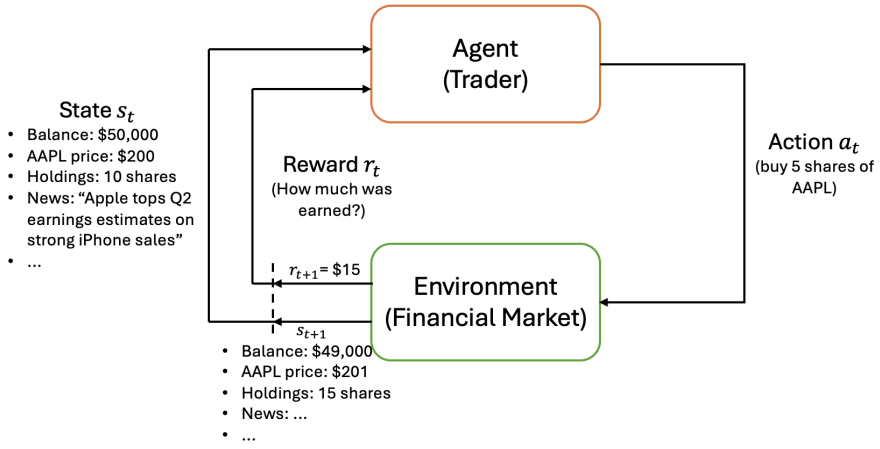


[6] Liu XY, Xia Z, Yang H, et al. Dynamic datasets and market environments for financial reinforcement learning. Machine Learning - Nature. 2024.

Financial reinforcement learning (FinRL) is an interdisciplinary field that applies RL algorithms to financial tasks, such as algorithmic trading, portfolio management, option pricing, hedging, and market making [1-5]. Developed by SecureFinAI Lab at Columbia University**,** FinRL aims to automate the design pipeline of a DRL trading strategy, allowing for efficiency, customization, and hands-on tutoring. One of the current research focuses on the integration of large language model (LLM)-generated signals from financial documents into FinRL, leveraging structured market data and unstructured textual data. From 2023 to 2025, we have organized three **FinRL contests** based on FinRL research. These contests cover diverse tasks such as stock trading, crypto trading, and LLM-generated signals for FinRL. The contests attracted around 200 students, researchers, and industry practitioners from over 100 institutions and 22 countries in total.

Why RL is a right language?

DRL is particularly well-suited for financial tasks because it can solve dynamic and sequential decision-making problems, which are very common in finance. For example, trading decisions are made continuously over time. Each trading decision depends on the current market conditions and the trader's portfolio. These decisions are sequential -- today's action affects tomorrow's opportunities, and feedback (e.g., gains or losses) helps refine future decisions.



This figure shows this process in FinRL with an example. The agent represents the trader for decision-making and the environment represents the financial market. The trader (agent) observes the market conditions (state s_t), such as its account information, market prices, and financial news. It also perceives the profit earned before, and anticipates the price of AAPL will rise. So, the trader (agent) decides to buy 5 shares of AAPL (action a_t). After taking this action, the market (environment) will update (state s_{t+1}), and the trader (agent) gets profits (reward r_{t+1} calculated as the change of total asset value). The trial-and-error learning allows the DRL agent to find a strategy that maximizes cumulative rewards over time, adapting to changing market conditions.

More formally, we formulate the trading task into a **Markov Decision Process**:

- **State** is a snapshot of the current market conditions. This is the information that traders know when making a decision. It typically includes account balance, asset prices, and holding shares, technical indicators (e.g., MACD, RSI), fundamental indicators (e.g., P/E ratio), LLM-generated signals from financial documents (e.g., sentiment signals of news), and so on.
- **Action** is what the agent can do as a trader. It can be buy, sell, and hold; or short and long; or adjust portfolio weights.
- **Reward** is an incentive signal. A common choice is the change in total asset value. It can also be a risk-adjusted return, such as the Sharpe ratio. The agent will learn how to maximize the positive reward and minimize the negative reward.

Through repeated interaction with the market environment, the agent learns a policy (typically a deep neural network) —a probability distribution over actions at state—that aims to maximize cumulative rewards over time.

Financial Data and Market Environment

The financial market data is dynamic, non-stationary, and has a low signal-to-noise ratio. FinRL transforms the raw data into standard market environments, where agents can learn and trade.

The financial data utilized in FinRL include:

- **OHLCV data**. OHLCV (open, high, low, close, volume) data is typical historical volume-price data in finance. This dataset is prepared through yfinance and FinRL-Meta. We also provide data APIs for various financial markets, held on the open-source repository FinRL-Meta.
- **Limit order book (LOB)** for cryptocurrency trading. LOB data offers a detailed view of market depth and liquidity, capturing the behavior of market participants and providing valuable insights into market trends. We use second-level LOB data of Bitcoin for crypto trading.
- **Financial news**. We utilize the Financial News and Stock Price Integration Dataset (FNSPID) [8], which contains 15 million time-aligned financial news articles from 1999 to 2023 for constituent companies in the S&P500. We also provide data APIs to access financial news and documents, held on the open-source repository FinGPT.

FinRL also provides the following features:

- **Market indicators**. The classical market indicators are calculated based on OHLCV data, including Moving Average Convergence Divergence, Bollinger Bands (upper and lower), Relative Strength Index, Commodity Channel Index, Directional Movement Index, Simple Moving Average for 30 days and 60 days, Volatility Index, and Turbulence.
- **ML-learned factors**. We adapted 101 formulaic alphas for LOB data and trained a recurrent neural network (RNN) to extract 8 strong factors. These factors have strong predictive power and capture complex market dynamics.
- **LLM-generated signals**. We utilize LLMs, such as DeepSeek-V3, to analyze financial news and generate actionable trading signals. An LLM assigns a sentiment score of 1 (strongly negative) to 5 (strongly positive) according to the news. It also assigns a risk level of 1(low risk) to 5 (high risk). These signals can then be included in the state and used to adjust the action and reward.

FinRL processes the market data and features into gym-style environments. For example, for trading 30 constituent stocks in Dow Jones Index:

- **State** is a 241-dimensional vector, including account balance, prices of 30 stocks, holding shares of 30 stocks, and 6 technical indicators (MACD, RSI, CCI, ADX, LLM-generated sentiment score, and LLM-generated risk level) for each of the 30 stocks.
- **Action** is a 30-dimensional vector. Each positive (negative) entry represents buying (selling) a certain amount of shares for that stock. The action is adjusted by the LLM-generated sentiment score – the action for a stock is amplified by up to 10% under positive sentiment and dampened by up to 10% under negative sentiment.
- **Reward** is the change in total asset value. To incorporate risk adjustment, LLM-generated risk levels—assigned to each stock based on news content—are aggregated into a weighted average portfolio-level risk factor. This factor is then used to penalize the reward when the overall risk level is high.

The state, action, and reward can be specified for different tasks. For example, the state can include more technical indicators, as stated above. The action can be continuous or discrete decisions. It can also support long-short trading strategies or be defined as portfolio weights in portfolio management tasks. The reward can use the risk-adjusted return Sharpe ratio in addition to the change in total asset value.

We also incorporate near-real market constraints in the environment:

- **Transaction costs**. A cost of 0.1% for each action {buy, sell} is set, accounting for commission and slippage.
- **Market volatility**. The Turbulence Index and VIX are risk indicators. Larger values indicate increased volatility due to investor fear and increased uncertainty, while smaller values indicate increased market stability.

FinRL also provides GPU-optimized massively parallel environments for simulation, which improves the sampling speed and addresses the sampling bottleneck efficiently. The experiment shows that with 2,048 parallel environments, the sampling speed can be improved by 1,650 times.

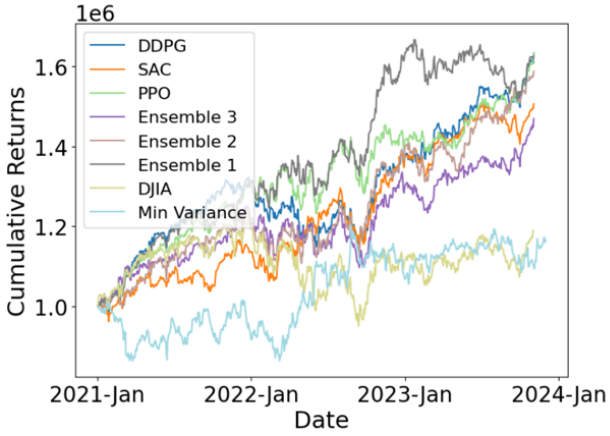
FinRL Agents: Learning to Trade

FinRL agents typically use a simple Multi-Layer Perceptron (MLP) as the policy network, which is a probability distribution over actions at state. FinRL supports more than 10 popular DRL algorithms, such as Proximal Policy Optimization (PPO), Advantage Actor-Critic (A2C), and Deep Deterministic Policy Gradient (DDPG), held in ElegantRL.

Evaluation and Performance

We evaluate models using well-established quantitative metrics in finance, such as cumulative return, Sharpe ratio, and maximum drawdown. These metrics provide a balanced view of performance and risk.

Here we show the experiment result of the stock trading task for 30 constituent stocks in the Dow Jones index. We use historical daily OHLCV data from 01/01/2021 to 12/01/2023 (734 trading days) and 10 market indicators listed previously. All models are trained, validated, and tested on a rolling-window basis with 30-day training, 5-day validation, and 5-day testing windows. We use PPO, SAC, and DDPG agents. The policy network for each agent consists of an MLP network with two hidden layers, having 64 units and 32 units, respectively. We set a learning rate of 3×10^{-4} and a batch size of 64. We also used ensemble methods to reduce policy instability. Ensemble 1 consists of 1 PPO, 1 SAC, and 1 DDPG agents; Ensemble 2 consists of 5 for each type; Ensemble 3 consists of 10 agents for each type. We use a weighted average approach to determine the final action. The performance is shown below. The PPO agent achieves the highest cumulative returns of 63.37%, Sharpe ratio of 1.55, and Sortino ratio of 2.44, showing an ability to maintain high returns with controlled volatility and downside risk. Ensemble 1 shows superior performance from Sep 2022 to Oct 2023. It also achieves the smallest maximum drawdown. All individual agents and ensemble models significantly outperform two traditional baselines, DJIA and min-variance strategy, across all metrics.



Model	Ensemble-1	Ensemble-2	Ensemble-3	PPO	SAC	DDPG	Min-Variance	DJIA
Cumulative Return	62.60%	58.77%	46.89%	**63.37%**	50.62%	63.19%	13.9%%	18.95%
Annual Return	18.22%	17.25%	14.15%	**18.41 %**	15.14%	18.36%	7.34%	6.15%
Annual Volatility	11.76%	12.61%	12.70%	**11.35%**	11.67%	11.93%	18.16%	15.14%
Sharpe Ratio	1.48	1.33	1.11	**1.55**	1.27	1.47	0.48	0.47
Sortino Ratio	2.34	2.14	1.74	**2.44**	2.05	2.37	0.73	0.67
Max Drawdown	**-8.98%**	-11.27%	-12.27%	-9.96%	-12.02%	-13.15%	-14.9%	-21.94%
RoMaD	**6.97**	5.22	3.82	6.36	4.21	4.81	1.10	0.86
Calmar Ratio	**2.03**	1.53	1.15	1.85	1.26	1.40	0.49	0.28
Omega Ratio	1.31	1.28	1.23	**1.33**	1.27	1.32	1.09	1.08

[7] Holzer N, Wang K, Xiao K, Yanglet XYL. Revisiting Ensemble Methods for Stock Trading and Crypto Trading Tasks at ACM ICAIF FinRL Contest 2023-2024. arXiv:2501.10709. 2025.

Conclusion

FinRL has already shown its potential in a range of applications, such as algorithmic trading, portfolio management, and option pricing. With its ability to incorporate signals from market data and financial news, FinRL allows agents to make more informed decisions.

Looking ahead, FinRL will continue exploring and integrating LLM-generated signals from multimodal financial data, such as SEC filings, earnings conference calls, alternative data, etc. This will build agents that better reflect how professional investors synthesize information.

Keyi Wang, SecureFinAI Lab at Columbia University, and The Beryl Consulting Group collaborated on this report.

References

[1] Liu XY, Yang H, Chen Q, et al. FinRL: A deep reinforcement learning library for automated stock trading in quantitative finance. Deep Reinforcement Leanring Workshop, NeurIPS. 2020.

[2] Liu XY, Yang H, Gao J, Wang CD. FinRL: deep reinforcement learning framework to automate trading in quantitative finance. ACM International Conference on AI in Finance. 2022.

[3] Hambly B, Xu R, Yang H. Recent advances in reinforcement learning in finance. Mathematical Finance. 2023;33(3):437–503.

[4] Sun S, Wang R, An B. Reinforcement learning for quantitative trading. ACM Transactions on Intelligent Systems and Technology. 2023;14(3):1–29

[5] Bai Y, Gao Y, Wan R, Zhang S, Song R. A review of reinforcement learning in financial applications. Annual Review of Statistics and Its Application. 2025;12(1):209–232.

[6] Liu XY, Xia Z, Yang H, et al. Dynamic datasets and market environments for financial reinforcement learning. Machine Learning - Nature. 2024.

[7] Holzer N, Wang K, Xiao K, Yanglet XYL. Revisiting Ensemble Methods for Stock Trading and Crypto Trading Tasks at ACM ICAIF FinRL Contest 2023-2024. arXiv:2501.10709. 2025.