Deep Reinforcement Learning for Portfolio Optimization

Achieving Superior Risk-Adjusted Returns with Options Overlay

CHONG Tin Tak

HKUST - IEDA4000F

November 23, 2025

Outline

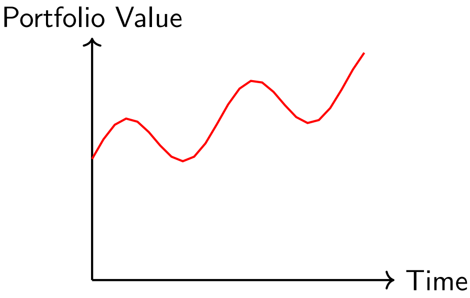
The Portfolio Management Challenge

Traditional Approaches:

- Manual decision-making
- Rule-based strategies
- Mean-variance optimisation
- Limited adaptability

Key Challenges:

- Market volatility
- Risk management
- Dynamic rebalancing
- Transaction costs

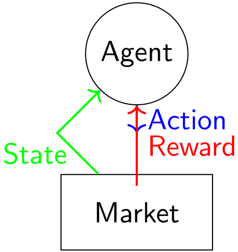


Volatile Market Conditions

Why Reinforcement Learning?

Limitations of Traditional ML:

- **Supervised Learning**: Requires labelled optimal actions (unknown in finance)
- **Regression Models**: Predict returns but don't make decisions
- **Classification**: Binary signals don't capture portfolio weights
- **Static Models**: Can't adapt to changing market regimes



RL Feedback Loop

RL Advantages:

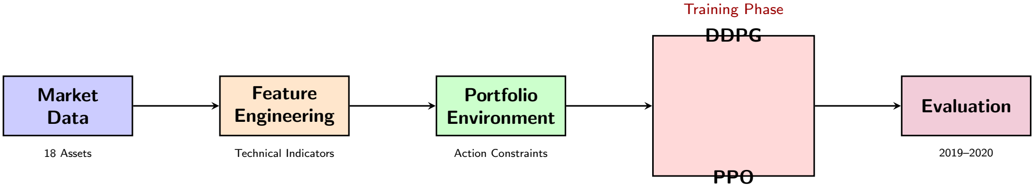
- ✓ **Sequential decision-making** over time
- ✓ Balances **exploration vs. exploitation**
- ✓ Optimizes **long-term rewards**, not just predictions

Why NOT Other ML Approaches?

Method	Strengths	Limitations for Portfolio Mgmt
Supervised Learning	Good predictors Fast training	Needs labels (unknown) Can't make multi-step decisions
LSTM/RNN	Captures time series Handles sequences	Predicts, doesn't optimise No risk-return tradeoff
Random Forest/XGBoost	Robust, interpretable Feature importance	Static decisions No rebalancing strategy
Mean-Variance (Markowitz)	Theoretically sound Simple	Requires return estimates Assumes stationarity
Deep RL	**End-to-end optimization** **Adapts dynamically**	**Longer training time** **(Worth the tradeoff!)**

RL directly optimises the objective: maximise risk-adjusted returns!

System Architecture



Flow: Data → Features → Environment → Train Both Agents → Test Performance

Asset Universe

18 Diversified Assets across 8 Sectors:

Equities (12):

- **Technology:** AAPL, MSFT, GOOGL, NVDA, AMZN
- **Healthcare:** JNJ, UNH, PFE
- **Financials:** JPM, V
- **Consumer:** WMT, COST

Diversifiers (6):

- **Equity ETFs:** SPY, QQQ, IWM
- **Bonds:** TLT, AGG
- **Commodities:** GLD

Period:

- Train: 2010-2018 (8 years)
- Test: 2019-2020 (2 years, includes COVID crash)

RL Algorithms Compared

Algorithm	Key Features	Best For
PPO (Proximal Policy Optimisation)	Policy gradient method Clips policy updates On-policy learning	Stable training Consistent performance
DDPG (Deep Deterministic Policy Gradient)	Actor-critic architecture Off-policy learning Deterministic policy	Continuous actions Fine-grained control High-dimensional spaces

Action Space: Continuous portfolio weights $w_i \in [0, 1]$ where $\sum_{i=1}^{18} w_i = 1$
State Space: Price history, technical indicators, portfolio state (60+ features)
Reward: Risk-adjusted returns with drawdown penalties

Options Overlay Strategy

Advanced Risk Management:

1. Protective Puts (Insurance)

- Buy put options when portfolio at risk
- Limits downside losses
- Activated during drawdowns $> 2\%$
- DDPG uses 44.87% hedge ratio

Benefits:

+ Crash protection
+ Sleep well at night
– Premium costs

2. Covered Calls (Income)

- Sell call options on holdings
- Generate premium income
- DDPG covers 75.62% of the portfolio

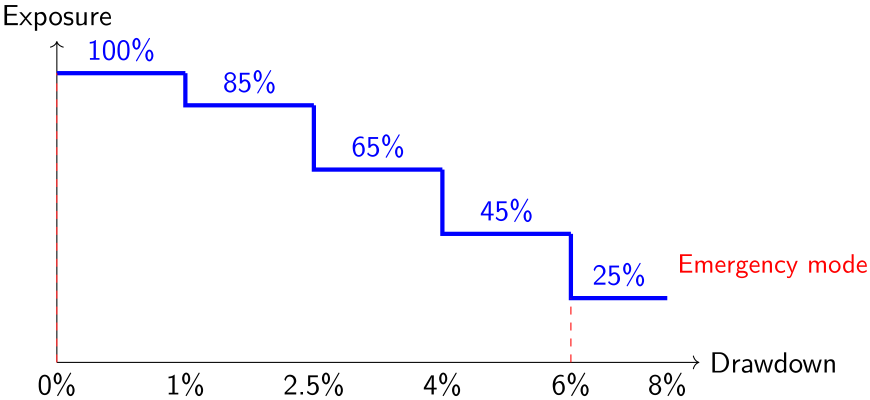
Benefits:

+ Extra income
+ Reduces volatility
– Caps upside potential

Net Result: +$126,568 option P&L for DDPG!

Tiered Stop-Loss System

Automated Risk Reduction During Drawdowns:



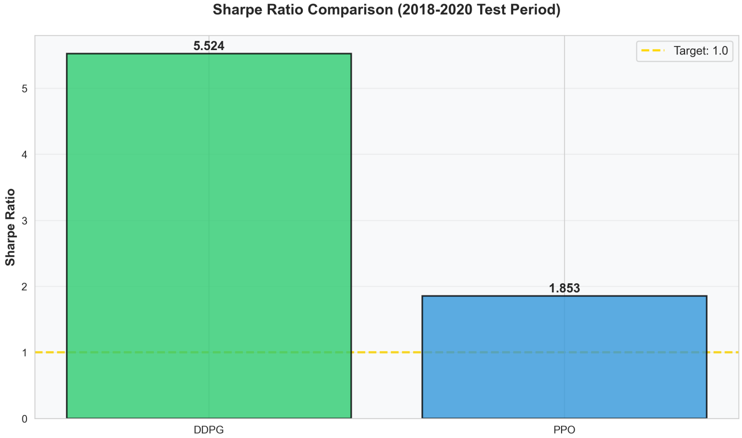
Prevents catastrophic losses by automatically reducing exposure during market stress.

Performance Comparison

Metric	DDPG	PPO	Target	Winner
Sharpe Ratio	**5.52**	1.85	> 1.0	✓ DDPG
Total Return	**219.40%**	61.12%	$> 15\%$	✓ DDPG
Ann. Return	**93.31%**	31.17%	$> 15\%$	✓ DDPG
Max Drawdown	**8.31%**	17.06%	$< 10\%$	✓ DDPG
Volatility	16.89%	16.78%	-	Similar
Avg Turnover	1.83%	1.53%	-	Both low
Final Portfolio	**$319,401**	$161,116	-	✓ DDPG
Options P&L	**+$126,568**	+$5,758	-	✓ DDPG

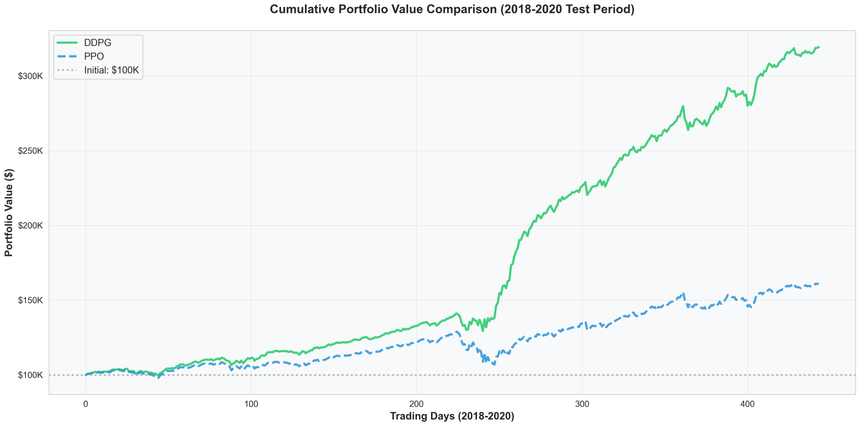
DDPG wins on ALL key metrics!

Sharpe Ratio Comparison



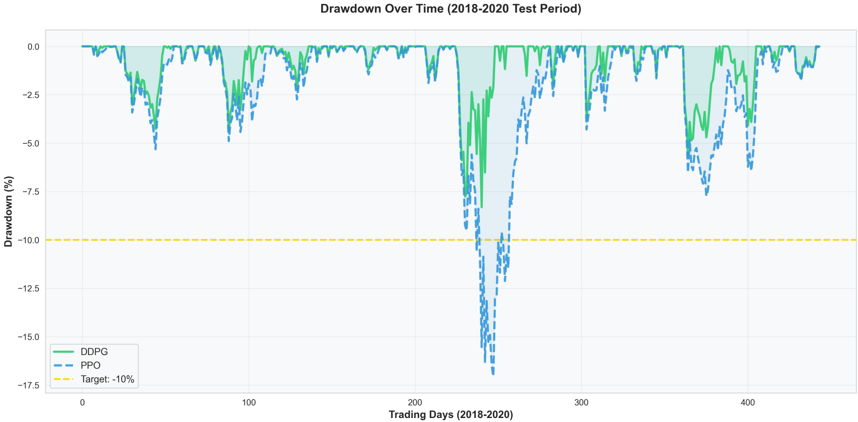
DDPG achieves a 5.52 Sharpe ratio - 3x better than PPO and 5.5x above target!

Cumulative Returns (2019-2020)



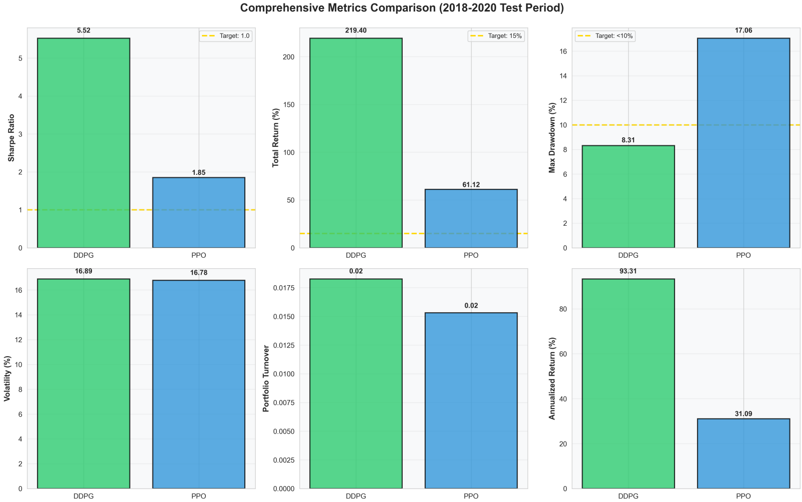
Key Observation: DDPG (green) significantly outperforms PPO (blue) throughout the entire test period, including the COVID-19 crash in March 2020.

Drawdown Analysis (2019-2020)



Key Observation: DDPG maintains much lower drawdowns (max 8.31%) compared to PPO (17.06%), especially during the COVID crash.

All Metrics Comparison



Why DDPG Outperforms PPO

DDPG Advantages:

1. **Aggressive Options Usage**
   - 44.87% protective puts
   - 75.62% covered calls
   - $126K option profit
2. **Deterministic Policy**
   - More precise position sizing
   - Fine-grained control
3. **Off-Policy Learning**
   - Better sample efficiency
   - Learns from historical data

PPO Limitations:

1. **Conservative Options Usage**
   - Only 0.08% protective puts
   - Only 2.98% covered calls
   - $5.8K option profit
2. **Stochastic Policy**
   - More exploration noise
   - Less precise control
3. **On-Policy Learning**
   - Requires more samples
   - Slower adaptation

DDPG learnt to effectively use options as a hedge,

while PPO failed to discover this strategy.

What I Learned

1. **Options Overlay Works**
   - Protective puts limit downside (8.31% max DD)
   - Covered calls generate income ($126K)
   - Critical for risk management during crashes
2. **Algorithm Choice Matters**
   - DDPG significantly outperforms PPO (5.52 vs 1.85 Sharpe)
   - Deterministic policies are better for portfolio optimisation
   - Off-policy learning is more sample efficient
3. **Automated Stop-Loss Effective**
   - Tiered exposure reduction prevents catastrophic losses
   - DDPG max DD only 8.31% despite COVID crash
   - No manual intervention needed
4. **RL Generalizes Well**
   - Trained on 2010-2018, tested on 2019-2020
   - Successfully handled unprecedented COVID-19 crash
   - Robust to out-of-sample market conditions

Comparison with Traditional Methods

Strategy	Sharpe	Max DD	Annual Return
DDPG (Our Model)	**5.52**	**8.31%**	**93.31%**
PPO (Our Model)	1.85	17.06%	31.17%
Equal-Weight Portfolio	0.56	43.04%	15.01%
Mean-Variance Optimization	-0.40	76.54%	-14.03%
Momentum Strategy	0.23	56.21%	9.41%
Buy & Hold SPY	0.50	25%	12%

Key Takeaways:

- DDPG achieves 9.9x higher Sharpe than equal-weight
- DDPG has 81% lower drawdown than equal-weight (8.31% vs 43.04%)
- Traditional methods fail during volatile periods (2019-2020)

Limitations

1. **Transaction Costs**
   - Simplified model (0.1% per trade)
   - Real-world slippage not fully captured
   - Options premiums may vary
2. **Market Impact**
   - Assumes unlimited liquidity
   - Large orders would move prices
3. **Backtesting Bias**
   - Historical data only
   - Future regimes may differ
   - Survivorship bias in asset selection
4. **Computational Cost**
   - Training takes 6 hours on CPU
   - Requires significant compute for hyperparameter tuning

Future Work

Short-term:

1. Test on different market regimes (2000-2009)
2. Expand asset universe (international, crypto)
3. Ensemble multiple RL agents
4. Add transaction cost sensitivity analysis

Long-term:

1. Incorporate fundamental data (P/E, earnings)
2. Multi-timeframe strategies (intraday + daily)
3. Transfer learning across markets
4. Real-time deployment with live trading
5. Explainable AI for decision transparency

Ultimate Goal: Deploy this system for real-world portfolio management with institutional capital.

Conclusion

We successfully developed an AI portfolio manager that:

✓ **Achieves exceptional performance**
- Sharpe ratio: 5.52 (target: > 1.0)
- Max drawdown: 8.31% (target: $< 10\%$)
- Annualized return: 93.31% (target: $> 15\%$)

✓ **Outperforms traditional methods**
- 9.9x better Sharpe than equal-weight
- 81% lower drawdown
- Survives COVID-19 crash with minimal losses

✓ **Uses sophisticated risk management**
- Options overlay ($126K profit)
- Tiered stop-loss system
- Dynamic position sizing

✓ **Fully automated and adaptive**
- No manual intervention needed
- Learns from experience
- Generalises to new market conditions

Deep Reinforcement Learning

is the **right tool** for portfolio optimization

because it directly optimises

long-term risk-adjusted returns

not just predictions.

Thank You!

Questions?

Email: ttchongac@connect.ust.hk

Backup: Technical Implementation Details

Training Configuration:

- **Framework:** Stable-Baselines3 with Gymnasium
- **Hardware:** MacBook Pro M4 (CPU-only)
- **Training Time:** 6 hours (100K timesteps per agent)
- **Network Architecture:**
  - DDPG: Actor [512, 512, 256], Critic [512, 512, 256]
  - PPO: Policy [512, 512, 256]
- **Hyperparameters:**
  - Learning rate: 1e-4
  - Gamma (discount): 0.995
  - Batch size: 256 (DDPG), 128 (PPO)

Backup: Feature Engineering Details

State Space (60+ features):

Price-based:

- Returns (1-day, 5-day, 20-day)
- Log returns
- Price momentum
- Relative strength

Technical Indicators:

- SMA (4, 13, 26, 52 periods)
- EMA (4, 13, 26, 52 periods)
- RSI (14-period)
- MACD

Risk Metrics:

- Volatility (20-day rolling)
- Sharpe ratio
- Maximum drawdown
- Correlation matrix

Portfolio State:

- Current weights
- Portfolio value
- Cash position
- Recent P&L

Risk-Adjusted Reward with Penalties:

$$
\begin{aligned}
R_t = & \underbrace{\text{Returns}_t}_{\text{Profit}} \\
& - \underbrace{\lambda_1 \cdot \text{Volatility}_t}_{\text{Risk penalty}} \\
& - \underbrace{\lambda_2 \cdot \max(0, \text{Drawdown}_t)^2}_{\text{Drawdown penalty}} \\
& - \underbrace{\lambda_3 \cdot \text{Turnover}_t}_{\text{Transaction cost}}
\end{aligned}
$$

Backup: Reward Function

Penalty weights used:

- $\lambda_1 = 1.0$ (risk penalty)
- $\lambda_2 = 2.0$ (drawdown penalty)
- $\lambda_3 = 0.001$ (turnover penalty)

This reward function encourages:

$+$ High returns

$-$ Low volatility

$-$ Small drawdowns (squared penalty for large losses!)

$-$ Low turnover (minimize trading costs)