RL Portfolio Training Loop

```
┌─────────────────────────────┐
│       Market Data           │
│ (Prices, Volumes, Factors)  │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│    Feature Engineering       │
│ (Returns, Indicators,        │
│  Volatility)                 │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│    Observation Vector        │
│         (o_t)                │
└─────────────────────────────┘
```

Policy / Value Networks
(Actor-Critic)

Softmax Allocation
(pi_t)

Portfolio Execution
& Environment

Next observation

Reward Function
(Sharpe, CVaR, etc.)

Gradients

Trajectory Buffer
/ Policy Update