Practitioner Brief

AI IN ASSET MANAGEMENT

Reinforcement Learning and Inverse Reinforcement Learning: A Practitioner's Guide for Investment Management

Igor Halperin, PhD, Petter N. Kolm, PhD, and Gordon Ritter, PhD

Practitioner Brief written by Mark Fortune

CFA Institute Research Foundation

CFA Institute Research & Policy Center

Finance is retooling decisions. Portfolio managers, risk chiefs, quants, and fintech builders are moving past static predictors that cannot keep up with faster, interconnected markets. Reinforcement learning (RL) and inverse reinforcement learning (IRL) offer action-first, trial-and-error strategies that adapt under uncertainty.

This chapter in *AI in Asset Management: Tools, Applications, and Frontiers* offers a practical guide to concepts, setup, training, and high-value quant uses. The chapter starts with the basics of decision making and then gets hands on: how modern, neural network RL works in practice—setup, training, data, costs, risks—and how to graduate from backtests to live implementation. It closes with demonstrated plays, including trade execution, portfolio rebalancing, hedging, and market making.

Who Should Read This Chapter?

This chapter targets those who design, run, or oversee investment decisions. Portfolio managers and traders can use it to turn forecasts into adaptive, cost-aware policies. Risk officers, compliance teams, and regulators gain clear ways to embed risk limits, auditability, and oversight. Quant researchers and data scientists get practical RL/IRL setups—from state design and rewards to "offline–sim–online deployment," which is a method for deploying and validating a system, especially in robotics and machine learning, that uses offline reinforcement learning and simulation before real-world online deployment. And fintech product leaders learn how to turn these methods into reliable, explainable tools.

Why This Chapter Matters Now

As markets move faster and change regimes more often, they can punish static, forecast-only models. RL and IRL allow practitioners to act—not just predict—by learning policies that adapt to costs, market impact, and delayed outcomes under uncertainty and partial information. They also build risk in from the start—for example, conditional value at risk (CVaR) and drawdowns—and fit finance's offline-data reality with sim-to-live deployment. As oversight tightens, these methods offer clearer objectives, audit trails, and guardrails—turning AI into accountable, real-world decision systems.

What Does This Chapter Deliver?

This chapter delivers a practical playbook for applying RL and IRL in finance. It explains step-by-step decision making (states, actions, rewards), contrasts RL with prediction models,

and guides algorithm choices (value based, policy gradient/actor-critic, model free versus model based; offline versus online). It shows how to make policies risk aware from the start (mean–variance, CVaR, distributional RL) and handle partial information. And it demonstrates concrete uses in execution, rebalancing, hedging, market making, surveillance, and consumer finance.

"As computational power grows and algorithms advance, RL and IRL are poised to become indispensable components of the quantitative finance toolkit, driving widespread adoption across the industry."

Igor Halperin, PhD, Petter N. Kolm, PhD, and Gordon Ritter, PhD

Practical Applications

Adaptive trade execution

- *Practitioners:* Use RL to slice large orders in real time based on liquidity, spreads, and volatility to cut implementation shortfall. Track slippage, market impact, and participation caps.
- *Policymakers:* Define "best-execution" evidence standards; require audit logs of policy decisions and cost attribution.

Dynamic, risk-aware portfolio rebalancing

- *Practitioners:* Train policies that rebalance across horizons while charging themselves for turnover/impact and targeting CVaR or drawdown. Monitor regime drift.
- *Policymakers:* Incorporate RL portfolios into stress tests; require documentation of risk objectives, constraints, and promotion criteria.

Cost-aware option hedging

- *Practitioners:* Replace naive delta hedging with RL/PPO (proximal policy optimization) that hedges discretely, respects lot sizes/fees, and minimizes tail loss. Measure hedging error and net P&L after costs.
- *Policymakers:* Set model risk expectations for discrete, costed hedging; require challenge models and scenario disclosures.

Liquidity provision/market making

- *Practitioners:* Train agents to set quotes that balance spread capture with inventory and adverse selection risk; enforce runtime inventory and loss limits.
- *Policymakers:* Monitor market quality metrics (spreads, depth, fill rates); use IRL to spot behaviors that amplify instability.

Market surveillance and abuse detection

- *Practitioners (venues):* Apply IRL to infer traders' objectives from order flow; flag patterns consistent with spoofing/layering before harm compounds.
- *Policymakers:* Use inferred-intent clustering to prioritize investigations, measure alert precision/recall, and require explainable triggers.

Consumer finance pricing and product design

- *Practitioners:* Use IRL to learn customer utilities from usage (e.g., data plans, credit), then run RL to propose fairer, better-fit plans; conduct A/B test with guardrails.
- *Policymakers:* Set transparency/fairness checks (opt in, bias audits, caps on penalty structures), and require outcome monitoring, not just model specs.

Practitioner Toolkit

The following provides a guide for how practitioners in key financial roles can apply reinforcement learning and inverse reinforcement learning.

Applications of RL and IRL by Role

Role	Key Techniques	Primary Applications	Main Benefits
Portfolio managers and traders	• Policy gradient/actor-critic (e.g., PPO) for continuous sizing • Distributional RL (e.g., QR-DQN/IQN) for tail control	• Adaptive trade execution beyond TWAP/VWAP (time-weighted average price and volume-weighted average price) • Dynamic portfolio rebalancing with turnover/impact costs	• Lower slippage and market impact • Higher risk-adjusted returns after costs
Risk and model risk/compliance	• Risk-sensitive rewards (mean–variance, CVaR, drawdown) • Policy constraints and safe-RL guardrails (train and runtime)	• Stress testing across regimes with challenger policies • Limit frameworks and policy approval/monitoring	• Transparent, auditable decision processes • Improved tail-risk control and resilience
Quant researchers and data scientists	• Markov decision process design: decision-time state, action, reward with costs • High-fidelity simulators and offline–sim–online pipeline	• End-to-end execution/hedging/allocation policies • Develop and validate market simulators (microstructure/latency)	• Better sample efficiency and out-of-sample robustness • Reduced model risk via validated simulators
Regulators, supervisors, and venues	• IRL (MaxEnt/Bayesian/AIRL/T-REX) to infer objectives • Behavior clustering and anomaly detection from order flow	• Detect spoofing/layering via inferred intent • Assess best-execution evidence and cost attribution	• Earlier, more precise detection • Evidence-based, explainable supervision
Fintech product leaders and engineers	• Policy-as-a-service with configurable objectives/constraints • SRE (site reliability engineering) for models: canaries, kill switches, drift monitors	• Execution/rebalancing products and robo-advice engines • Consumer pricing/plan recommendations with IRL and RL	• Faster time to value with safer deployments • Differentiated, explainable products for clients and regulators

Implementation

Integrate RL and IRL by layering them onto your current data, research, and execution stacks. Start offline, using historical prices, fills, and positions to define states, actions, and rewards that include costs and risk. Use IRL to infer underlying objectives and feed that reward into RL training. Validate in a high-fidelity simulator, and then run the policy in shadow mode alongside your existing process. Expose it via an application programming interface (API) to an order management system/execution management system or risk engine with hard limits, approvals, and a kill switch. Log state–action–outcome for audits, monitor drift and performance with dashboards, and promote versioned policies through your model governance workflow.

Metrics That Matter

From the many measures available, this chapter highlights three metrics as especially critical for practitioners.

- *Implementation shortfall (IS):* IS measures the real cost of execution versus the decision/benchmark price. It captures slippage, market impact, and timing risk, and the chapter even frames execution rewards as the *negative* of IS. Track per order/strategy and set caps.
- *Conditional value at risk:* CVaR is a tail-risk metric central to risk-aware RL. Use it as a penalty or constraint in training and as a live limit so policies do not buy returns by loading the left tail.
- *Risk-adjusted return after costs (e.g., Sharpe or mean–variance utility):* Optimize and report performance *net* of transaction/impact costs, balancing expected return against variance as recommended (and validated by the utility/mean–variance linkage).

Glossary

Inverse reinforcement learning: Works backward from observed behavior (e.g., trades) to infer the hidden objective or "reward" the expert is optimizing so that it can be modeled and improved.

Markov decision process: The formal setup for RL: states (what you know now), actions (what you can do), transition dynamics (what likely happens next), rewards (immediate feedback), and a discount factor.

Offline–sim–online (deployment pipeline): A staged rollout for decision models.

Policy (π): The strategy the agent uses: a mapping from state to action (deterministic or probabilistic). Training an RL system means learning a good policy.

Reinforcement learning: A method where an agent learns by trial and error to choose actions that maximize long-run reward in a changing market, accounting for costs, impact, and delayed effects.

Reward & Return (R, γ): The reward is the step-by-step feedback (e.g., net P&L minus costs and risk penalties). The return is the discounted sum of future rewards, using γ to value near-term versus later outcomes.

Related Content

Pisaneschi, Brian. 2025. "Agentic AI for Finance: Workflows, Tips, and Case Studies." CFA Institute Research and Policy Center. https://rpc.cfainstitute.org/research/the-automation-ahead-content-series/agentic-ai-for-finance.

Spyrou, Alex, and Brian Pisaneschi. 2024. "Practical Guide for LLMs in the Financial Industry." CFA Institute Research and Policy Center. https://rpc.cfainstitute.org/research/the-automation-ahead-content-series/practical-guide-for-llms-in-the-financial-industry.

Tait, James. 2025. "Synthetic Data in Investment Management." CFA Institute Research and Policy Center. https://rpc.cfainstitute.org/research/reports/2025/synthetic-data-in-investment-management.

Wilson, Cheryll-Ann. 2025. "Explainable AI in Finance: Addressing the Needs of Diverse Stakeholders." CFA Institute Research and Policy Center. https://rpc.cfainstitute.org/research/reports/2025/explainable-ai-in-finance.

CFA Institute Research Foundation

CFA Institute Research & Policy Center

Find *AI in Asset Management: Tools, Applications, and Frontiers* and more on the CFA Institute Research and Policy Center website: https://rpc.cfainstitute.org.

Read the full chapter