Autonomous AI Feature Learning and Signal Discovery for Quantitative Finance

Vision Summary We propose an advanced closed-loop system that combines state-of-the-art financial representation learning with autonomous, agentic GPT-5-level reasoning to discover, evaluate, and refine predictive features for trading strategies. At the core is a tight integration of: - Self-supervised embeddings (e.g., TS2Vec, DeepLOB, InceptionTime) - Symbolic regression and neuro-symbolic distillation (e.g., PySR) - Reinforcement learning for feature search - Regime-aware backtesting - Vector database for similarity-based pattern recall - A GPT-5-style planner-agent that reasons over strategy outcomes, rewrites feature pipelines, and orchestrates continual learning

This system is designed to not only discover alpha, but to *learn how to learn alpha*.

System Architecture

1. **Raw Data Pipeline**: Ingests tick, LOB, filings, dark pool, and macro data. Cleans and segments into context windows.

2. **Feature Generator Layer**

3. Deep: TS2Vec, DeepLOB, InceptionTime for latent embeddings
4. Symbolic: PySR/AI-Feynman for equation discovery
5. Topological: Persistent homology for regime topology

6. Change-point and regime models (BOCPD, HMMs)

7. **Vector Database Layer**

8. Stores embeddings and symbolic fingerprints
9. Enables similarity search: "When did we last see a signal like this?"

10. Used by both the model and the agent for analogical reasoning

11. **Backtesting & Evaluation**

12. Rolling out-of-sample backtests using signal-defined strategies
13. Risk-adjusted metrics (Sharpe, Sortino, drawdown, t-stats)

14. Regime-specific scorecards and performance decay monitoring

15. **Agentic Planning Layer (GPT-5 Agent)**

16. Reads feature logs, backtest diagnostics, and regime outcomes
17. Designs new feature transformations, model architectures, symbolic derivations
18. Issues tool calls to: retrain models, evolve features, prune the vector DB, re-index drifted embeddings

19. Learns via outcome-aware feedback: features that fail in backtest are deprecated or re-learned

20. **Retraining Loop**

21. GPT-triggered or schedule-driven updates to all components
22. Maintains rolling windows and validation for continual learning

---

Why This Advances the State of the Art This vision stays faithful to the foundational research on self-supervised embeddings, symbolic discovery, and vector-based signal retrieval. It augments those with: - Planning and tool use (via agentic GPT) - Self-rewriting pipelines - Context-aware feature prioritization - Regime-adaptive validation

The result is a quant system that doesn't just build features—it *thinks* about features. It reasons over failure, adapts to new regimes, generalizes across conditions, and curates its own evolving alpha library.

Applications and Use Cases - Intraday or daily alpha signal discovery - Insider trading pattern correlation via vector similarity - Systematic risk regime detection - Adaptive hedge overlays and stop logic - Meta-portfolio optimization based on feature clusters

Conclusion With a GPT-5-level agent embedded into a live-feedback, backtest-grounded feature discovery loop, this system becomes a self-reinforcing alpha machine—capable not only of learning from market data, but of *strategically evolving its own intelligence* to navigate the market's future complexity.