EMNLP 2025 Paper Findings

Bridging Human Expertise & AI in Financial Decision-Making

FinHEAR: A novel multi-agent framework integrating legendary investor wisdom, behavioral economics, and adaptive risk-aware reasoning.

Start Presentation

Jump to Simulation

Introduction & Motivation

Addressing the limitations of standard LLMs in dynamic financial environments.

Abstract

Financial decision-making demands temporal reasoning, adaptive risk assessment, and responsiveness to dynamic events. While LLMs show strong general reasoning, they fail to capture behavioral patterns like **loss aversion** and **expert reliance**.

FinHEAR is a multi-agent framework that orchestrates specialized agents to analyze trends, interpret events, and retrieve expert-informed precedents. Grounded in behavioral economics, it incorporates expert-guided retrieval and confidence-adjusted position sizing.

Key Contributions

- **Multi-Agent Framework:** Integrates Human Expertise and Adaptive Risk-aware temporal reasoning within an event-driven pipeline.
- **Behavioral Grounding:** Explicit mechanisms for expert-guided retrieval and risk-aware action modulation based on *Prospect Theory*.
- **Empirical Validation:** Consistently outperforms baselines (FinCon, PPO) in trend forecasting and decision-making tasks.

Theoretical Foundations

FinHEAR treats finance as a human behavioral challenge, not just a math problem. Each card explores the cognitive foundation

Information Asymmetry

Markets contain uneven information distributions. Retail investors often lack the "insider" view.

Solution: Bounded Rationality Emulation
Retrieves expert-informed cases to mitigate asymmetry using heuristic shortcuts found in professional expertise.

Prospect Theory

Humans feel the pain of a loss more acutely than the joy of an equivalent gain (Loss Aversion).

Solution: Adaptive Risk-Awareness
Adjusts position sizes using a **Skewed Beta Distribution**, being aggressive in gains but conservative in high-risk contexts.

Temporal Consistency

Financial decisions unfold over time; early errors can compound if not revised.

Solution: Feedback Loop
A Temporal Refinement Agent updates historical trends based on realized outcomes to limit error propagation.

Methodology & Formulations

Mathematical modeling of expert retrieval and risk-aware decision making.

1. Expert Retrieval

We retrieve the most relevant expert case x_{exp} using cosine similarity of context embeddings $\phi(\cdot)$.

$$T_{retrieved} = \arg\max_{x_{exp} \in M} \cos(\phi(current), \phi(x_{exp}))$$

2. Adaptive Risk (Prospect Theory)

We sample a sensitivity score ρ from a Scaled Beta distribution, skewed based on risk context (High vs Low).

$$\rho \sim \text{ScaledBeta}(\alpha_r, \beta_r; [a_r, b_r])$$

High Risk: Right-skewed $\text{Beta}(2, 5)$ *(Conservative)*
Low Risk: Left-skewed $\text{Beta}(5, 2)$ *(Aggressive)*

3. Position Scoring

The final position score w combines risk sensitivity $f(\rho)$, expert reliability $h(\alpha)$, and event similarity $g(\gamma)$.

$$w = F(\rho, \alpha, \gamma) = f(\rho) \cdot h(\alpha) \cdot g(\gamma)$$

4. Decision Logic

Actions are determined by thresholds δ_{low} and δ_{high} and market direction d.

$$\text{Action} = \begin{cases} \text{close}, & w > \delta_{high} \\ \text{hold}, & w < \delta_{low} \\ \text{long}, & d = \text{up and } \delta_{low} \leq w \leq \delta_{high} \\ \text{short}, & d = \text{down and } \delta_{low} \leq w \leq \delta_{high} \end{cases}$$

Encoded Human Expertise

Aligning AI agents with legendary investment philosophies from curated expert knowledge base.

Warren Buffett

VALUE INVESTING

Best Match: **AAPL** (30.81% CR)

"Economic moats, DCF analysis, margin of safety. Focus on intrinsic value and long-term competitive advantages."

> *"Interest rates are to asset prices what gravity is to the apple. When rates are high, there's a downward pull."*

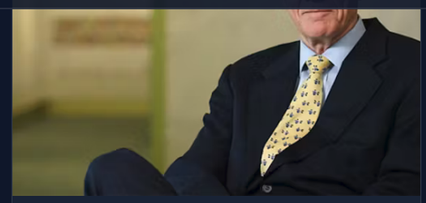
George Soros

REFLEXIVITY / MACRO

Best Match: **XAUUSD** (32.45% CR)

"Markets reflect and shape reality. Speculates on macroeconomic shifts, policy changes, and feedback loops."

> *"The financial markets generally are unpredictable. So one has to have different scenarios... The idea that you can actually predict what's going to happen contradicts my way of looking at the market."*



Peter Lynch

GROWTH AT REASONABLE PRICE

Best Match: **TSLA** (58.74% CR)

"Invest in what you know. Growth stocks with strong narratives, innovation catalysts, and scalable business models."

> *"When I see something I like, I buy it. I just hope it goes up before it goes down. The basic story remains favorable."*

Benjamin Graham

DEFENSIVE VALUE

Best Match: **XOM** (37.16% CR)

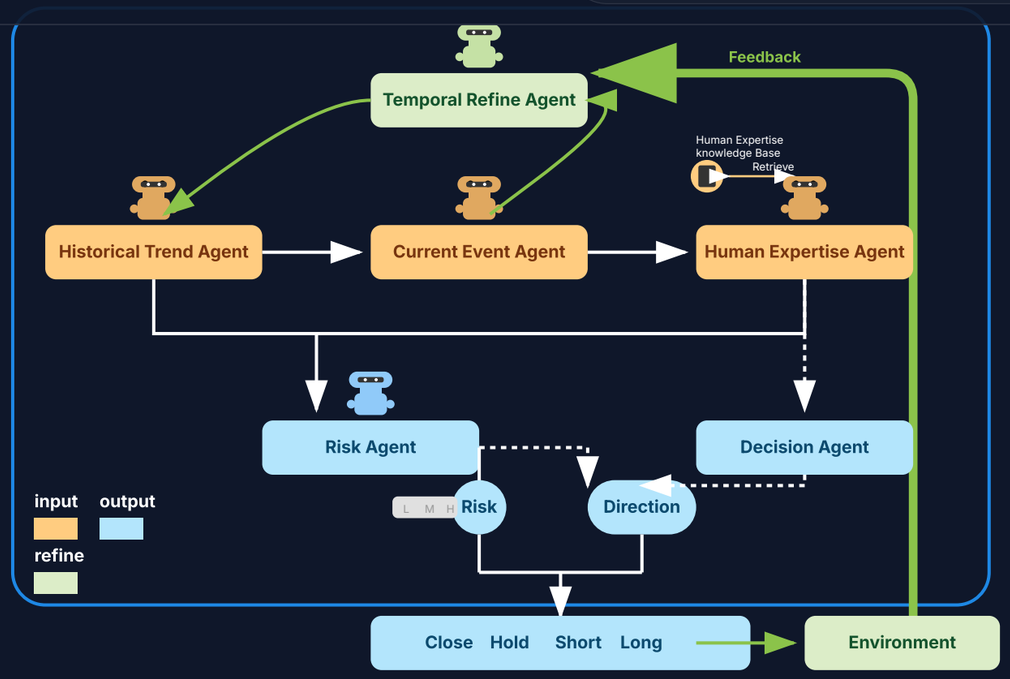
"Father of value investing. Focus on balance sheets, asset coverage, stable dividends, and safety margins."

> *"The intelligent investor is a realist who sells to optimists and buys from pessimists."*

Knowledge Construction: Expert cases extracted from books, letters, and interviews using GPT-4o, then matched via cosine similarity of embeddings (text-embedding-3-large). Successful predictions reinforce knowledge; failures trigger adaptive refinement.

System Architecture

A multi-agent framework orchestrating historical analysis, event interpretation, and expert retrieval.



Click on any agent to view its logic and prompt structure.

Interactive Model Walkthrough

Scenario: **AAPL** during a macro rate-hike scare.

Scenario 1: HOLD | **Scenario 2: BUY**

1. Inputs | 2. Expert | 3. Risk (Math) | 4. Decision

Market Signals

HISTORICAL TREND (AGENT 1)

Uptrend Detected

5-Day Momentum: +2.4%

CURRENT EVENT (AGENT 2)

Macro Headwind

Fed Minutes: "Rates may stay high."

MODEL INTERNALS VISUALIZATION

History +Buy VS Event -Sell

⚠ High Conflict Detected

THE MATHEMATICAL ENGINE

Input Vectors: H_t (History) and E_t (Event) show divergence.

$$\text{Conflict}(H_t, E_t) > \theta_{threshold} \implies \text{Trigger Risk Analysis}$$

Evaluation Metrics

FinHEAR uses rigorous financial and classification metrics to assess strategy effectiveness. *Click each metric to explore its mathematical foundation.*

Cumulative Return (CR)

Total profit/loss generated by the strategy, accounting for timing, direction, and position sizing.

Sharpe Ratio (SR)

Risk-adjusted return comparing profit above risk-free rate to volatility experienced.

Maximum Drawdown (MDD)

Largest peak-to-trough decline—quantifies worst-case portfolio loss.

Calmar Ratio

Return efficiency relative to worst drawdown—balances profit against maximum crash.

Accuracy (ACC)

Percentage of correct directional predictions (Up vs. Down).

Matthews Correlation Coefficient (MCC)

Balanced accuracy metric accounting for class imbalance—more robust than simple accuracy.

Performance Results

Comparative evaluation against state-of-the-art baselines.

Trend Forecasting | Trading Performance | Market Regimes

Trend Forecasting Accuracy

Accuracy (ACC%) and Matthews Correlation Coefficient (MCC) across four financial assets during the test period (Dec 2019 - Dec 2023).

CATEGORY	METHOD	XAUUSD		AAPL		TSLA		XOM	
		ACC ↑	MCC ↑	ACC ↑	MCC ↑	ACC ↑	MCC ↑	ACC ↑	MCC ↑
Rule-based	Momentum	47.58	-0.049	48.58	-0.037	51.82	0.026	47.18	-0.057
	Mean Reversion	47.35	-0.053	45.49	-0.074	47.76	-0.037	52.65	0.052
Deep-learning	ALSTM	54.32	0.121	54.69	0.174	54.88	0.110	52.25	0.059
	Informer	53.70	0.158	53.52	0.153	55.36	0.052	53.59	0.084
	StockNet	52.57	-0.037	53.59	0.087	55.77	0.042	51.96	0.049
	PPO	54.72	0.135	53.62	0.182	54.07	-0.009	56.52	0.124
LLM-based	CoT	52.03	0.149	56.13	0.163	51.83	0.062	51.27	0.059
	GA	54.47	0.155	54.17	0.108	56.54	0.141	52.78	0.062
	FinMem	55.73	0.142	55.32	0.133	58.82	0.133	53.91	0.043
	FinCon	56.81	0.141	52.76	0.071	57.46	0.119	56.37	0.143
Ours	**FinHEAR**	**58.97**	**0.202**	**59.08**	**0.196**	**61.56**	**0.190**	**58.19**	**0.163**

Ablation Studies

Component contribution analysis and expert-style alignment experiments.

Component Contribution Analysis

Trading performance of FinHEAR variants under single-asset (AAPL) and portfolio management tasks.

TASK	ASSET	ABLATION VARIANT	CR (%) ↑	SR ↑	MDD (%) ↓	CALMARR ↑
Single-Asset	AAPL	w/o temporal refinement	22.39	1.67	25.58	0.88
		w/o past trend	14.41	0.86	36.96	0.39
		w/o human expertise	19.75	0.32	35.15	0.56
		w/o adaptive risk-aware	16.32	0.74	31.72	0.51
		FinHEAR	**30.81**	**1.95**	**23.88**	**1.29**
Portfolio Mgmt.	XAUUSD, AAPL, TSLA, XOM	w/o temporal refinement	52.73	2.17	37.81	1.39
		w/o past trend	37.29	1.34	33.65	1.11
		w/o human expertise	42.12	1.22	28.94	1.46
		w/o adaptive risk-aware	36.18	1.15	35.03	1.03
		FinHEAR	**83.89**	**3.28**	**25.77**	**3.22**

Conclusion

FinHEAR bridges the gap between quantitative rigor and qualitative reasoning.

✅ Strengths

- First multi-agent framework combining **behavioral economics** with LLM reasoning
- Consistent **best-in-class performance** across all **tested** assets and metrics
- Robust in both **bull and bear markets**

⚠️ Limitations

- **High API costs** from GPT-4o dependency
- Tested on **only 4 assets** over a relatively short period
- Manual expert-to-asset assignment

🔮 Future Work

- **Adaptive expert selection** based on market context
- Fine-tune **smaller LLMs** for cost-efficient deployment
- Extended testing across **longer time horizons**

Read the Paper | View Code

© 2025 FinHEAR Research Team. EMNLP 2025.