RetailAgent: Auditing Systematic Timing Errors in Self-Conditioned LLM Agents

Anonymous authors
Paper under double-blind review

Abstract

Sequential LLM agents can fail in ways that are directional rather than noise-like. We introduce *RetailAgent*, a controlled multimodal, state-conditioned framework that repeatedly chooses long or flat from an anonymized intraday price history, its chart representation, and permitted account state. RetailAgent provides a sparse price-and-state interface for auditing sequential decisions; it neither represents a retail population nor simulates a full market. We evaluate each decision trajectory using exposure-matched within-stock timing, which tests alignment with subsequently revealed return labels rather than executable portfolio performance. Across text, chart, joint text–chart, and account-state conditions, all estimates are negative; the principal 10-minute text, chart, and multimodal arms are −45.7, −29.9, and −48.9 bps per stock-day. This wrong-signed timing is directional: the complementary schedule reverses its score, while linear projection on conventional price models leaves substantial residual timing. Narrative self-conditioning further reorganizes the trajectory: stronger wrong-signed timing coincides with fewer position changes in the tested memo conditions. We interpret these findings as structured sequential failure and as motivation for future counter-policy and leader–follower analyses, not as evidence of profitable trading, market impact, or human retail equivalence.

1 Introduction

Financial-market theory shows that prices can aggregate dispersed information and that order flow can reveal information to other market participants (Grossman & Stiglitz, 1980; Kyle, 1985). A policy that responds systematically to salient price movements can therefore be anticipated even when each action sounds plausible in isolation. This concern is especially relevant to individual investors: brokerage evidence links retail trading to limited attention, past returns, overconfidence, and costly turnover (Barber & Odean, 2000; 2008; 2013), and stocks favored by small traders have displayed subsequent underperformance in historical samples (Hvidkjaer, 2008). These findings do not imply that all retail investors share one policy. Rather, the top panel of Figure 1 provides a conceptual motivation: systematic price-conditioned reactions can create predictable behavior that may be informative to other market participants.

LLMs offer a controlled experimental object for isolating such a decision policy. LLM

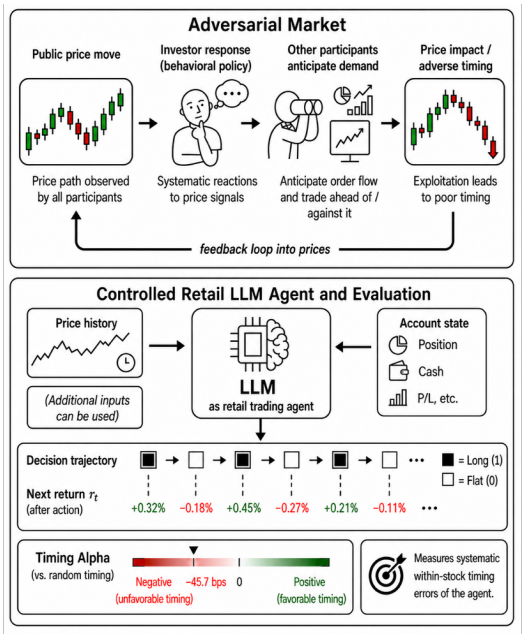


Figure 1: Why systematic sequential behavior matters. Top: predictable price-conditioned behavior may invite a counter-response in an interactive market. Bottom: RetailAgent isolates one such decision policy.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

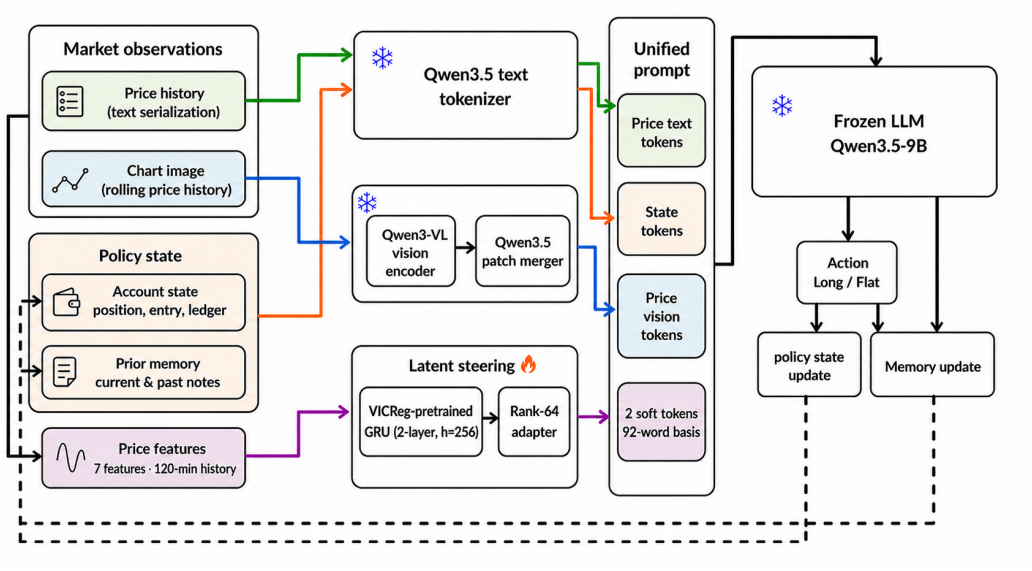


Figure 2: **RetailAgent architecture.** Price history and permitted account state use the fixed Qwen3.5 text tokenizer, while charts use the fixed Qwen3-VL vision encoder and patch merger. Prior memory is updated recursively; seven price features derived from the observed history pass through a VICReg-pretrained two-layer GRU and rank-64 adapter into two optional soft tokens on a 92-word basis. The unified prompt combines these representations before the frozen LLM chooses long or flat. The snowflake and flame distinguish the fixed backbone from the learned latent-steering branch. The subsequent return is withheld until after the action.

agents increasingly make sequential decisions while conditioning on previous actions, stored memories, and self-generated interpretations (Park et al., 2023; Liu et al., 2024); prior work has also used them as simulated economic agents and to reproduce aspects of human samples and experimental responses (Horton et al., 2023; Aher et al., 2023; Argyle et al., 2023). We operationalize *RetailAgent* as a controlled multimodal and state-conditioned proxy for one price-conditioned policy. At each interval, it receives an anonymized rolling intraday price history as text, chart, or both, together with permitted endogenous account state, and selects long or flat; the next return label is revealed only after the decision. Stock identity, date, absolute price, news, fundamentals, cross-sectional ranks, and future outcomes remain hidden. RetailAgent is neither a representative retail investor nor a model of the full adversarial market: it does not affect the historical price path or simulate strategic counterparties, endogenous order flow, or market impact. This fixed information boundary yields sequential observations, binary actions, and objectively realized subsequent outcomes, while separating within-stock timing from contemporaneous cross-sectional selection. Figure 2 details the fixed text and vision pathways, recursive policy state, and optional latent steering used by the audit.

With this controlled agent in hand, the central question is whether sequential LLM failures are merely noise-like or exhibit stable directional structure. Within-stock timing alpha asks whether the agent is long during relatively favorable subsequently realized intervals and flat during relatively unfavorable ones, after removing average exposure. Positive timing alpha indicates alignment, near-zero timing alpha is compatible with noise-like timing, and negative timing alpha indicates systematically wrong-signed timing. A consistently wrong agent need not be an uninformative agent. Across all conditions, RetailAgent has a negative within-stock timing estimate. The principal 10-minute Qwen arms are -45.7 bps per stock-day for text, -29.9 for chart, and -48.9 for text plus chart. We emphasize repeated sign robustness across representations, horizons, endogenous state, and model checks rather than ranking magnitudes from unmatched samples. Negative timing is informative in a way that a near-zero score is not: complementing every long/flat decision reverses the sign of timing alpha. Thus inversion leaves a near-zero signal near zero but reverses a wrong-signed one.

The existing projection analysis further shows that the tested GRU, GBDT, and reversal signals do not linearly account for the full inverted schedule.

Self-conditioning supplies a second finding. The agent can write an interpretation for its future self, making generated language part of its next decision context. Under the tested consistency conditions, moving from the current note only ($w = 0$) to one earlier note ($w = 1$) changes timing from -62.8 to -74.1 bps while position changes fall from 2.5 to 1.7 per stock-day. Thus self-conditioning increases persistence and coincides with stronger wrong-signed timing in these conditions; persistence and self-reference remain mechanistic correlates, not established mediation. More broadly, systematic behavior of one agent may define a response against which another decision maker could optimize. Writing $p = \pi_{\text{LLM}}(x, m)$, a hypothetical counter-agent could seek $\psi^\star \in \arg\max_\psi J_{\text{counter}}(\psi; \pi_{\text{LLM}})$. This is a leader–follower, bilevel-inspired perspective, not a contribution of the present method: we characterize the follower-side behavior and inspect the simplest complementary schedule; learning an adaptive counter-policy is future work.

Our contributions are as follows.

- **RetailAgent audit framework.** We formulate a controlled multimodal and state-conditioned interface for instantiating and auditing sequential LLM decision policies under a fixed information boundary (Section 3).
- **Structured adverse timing.** We show negative within-stock timing estimates across all 14 tested standard configurations, spanning representation, decision horizon, endogenous state, and model variation; the complementary-schedule identity and signal-overlap analysis make the directional structure explicit (Sections 4.2 and 4.3).
- **Conditioning and persistence.** We show that narrative self-conditioning substantially reorganizes temporal behavior: stronger wrong-signed timing coincides with fewer position changes, while latent conditioning provides a distinct state-based comparison channel (Section 4.4 and Appendix B).

The resulting claim is behavioral rather than economic: RetailAgent exposes structured sequential failure under a controlled interface. A deployable counter-policy or trading strategy would require executable prices, costs, capacity analysis, and a globally untouched market test.

2 Related work

Market behavior as motivation. Individual investors have historically underperformed market benchmarks after costs, with frequent trading accounting for part of the gap (Barber & Odean, 2000). Attention affects which securities individuals buy (Barber & Odean, 2008), while display format can change retail behavior and performance even when information content is held fixed (Benamar, 2019). Past gains can increase confidence, volume, and risk-taking even when they contain little information about future skill (Statman et al., 2006; **?**). These findings motivate a sparse price-and-state decision interface; they do not identify a representative retail policy, observe retail orders in our data, or explain RetailAgent's measured timing.

Language models in finance. Financial LLM research spans domain pretraining, financial reasoning, return prediction, and autonomous agents. BloombergGPT studies large-scale financial domain pretraining (Wu et al., 2023); FinGPT emphasizes open data pipelines and parameter-efficient adaptation (Liu et al., 2023). Recent benchmarks broaden financial evaluation from quantitative reasoning (Krumdick et al., 2024) to multi-task financial capabilities (Xie et al., 2024b), data analysis (Liu et al., 2025b), and multilingual multimodal applications (Peng et al., 2026). At the systems level, FinMem uses layered memory and character design (Yu et al., 2024a); FinCon uses a manager–analyst hierarchy and verbal reinforcement (Yu et al., 2024b); TradingAgents organizes specialized roles into a simulated trading firm (Xiao et al., 2024); and CryptoTrade combines on-chain and off-chain information with reflection (Li et al., 2024). These works seek stronger financial capabilities or investment decisions through added information, tools, or deliberative structure. RetailAgent instead

fixes a sparse information boundary and uses the market environment to measure structured sequential error.

LLM agents and behavioral simulators. Recent work has developed LLM agents that store, retrieve, and reflect on experience to produce behavior in interactive environments (Park et al., 2023), and has evaluated their decision making across heterogeneous tasks (Liu et al., 2024). Behavioral-simulation studies examine controlled experimental responses (Aher et al., 2023; Argyle et al., 2023), simulated economic agents (**?**), trust in an economic game (Xie et al., 2024a), and cooperation in a commons setting (Piatti et al., 2024). More recent evidence finds that persona prompting yields limited gains when persona variables explain little outcome variance (Hu & Collier, 2024), that models systematically overestimate human rationality (Liu et al., 2025a), and that broad simulation fidelity remains modest on a standardized benchmark (Hu et al., 2026). RetailAgent likewise treats the model as an experimental object, but asks whether a sequential policy's errors are directionally structured under a fixed information set rather than whether the model reproduces people or maximizes task reward.

Continuous conditioning and self-conditioning. Prefix tuning and prompt tuning adapt frozen language models by learning a small number of continuous embeddings (Li & Liang, 2021; Lester et al., 2021); analogous methods extend to multimodal models (Yang et al., 2023). Our steering module differs in that its soft tokens are input-dependent outputs of a self-supervised time-series encoder. The memo intervention is related to the memory and reflection mechanisms used in LLM-agent architectures (Park et al., 2023), but isolates a minimal mechanism: the model writes one note and then reads it before making the next decision, without parameter updates or new market information.

3 Methods

3.1 Controlled sequential decisions and evaluation

RetailAgent separates the market information available to an agent from the endogenous state through which it interprets that information. The experimental unit is a stock-day pair (i, d): one anonymized equity observed over one intraday session. At decision time t, the agent receives $x_{id,1:t}$, the rolling price history available at the beginning of that interval, and chooses a binary position $p_{idt} \in \{0, 1\}$, where 1 denotes long and 0 denotes flat. The subsequent interval has return label r_{idt}; this label is revealed only after the action and is not an executable quote or transaction-cost-adjusted return. Stock identity, calendar date, absolute price level, news, fundamentals, cross-sectional ranks, and future outcomes are concealed. Consequently, differences in p_{idt} across conditions can be attributed to changes in how the same permitted information is represented or carried forward, rather than to an expanded market information set.

For a stock-day with T decision intervals and return vector $r_{id} = (r_{id1}, \ldots, r_{idT})$, we measure exposure-matched timing alpha as

$$A_{id}(p, r) = \sum_t (p_{idt} - \bar{p}_{id})\, r_{idt}, \qquad \bar{p}_{id} = T^{-1} \sum_t p_{idt}. \tag{1}$$

This is a within-stock covariance score: it asks whether the agent is long during relatively favorable intervals after its average exposure is removed. We report the mean in basis points per stock-day. Positive values indicate directionally aligned timing, values near zero are compatible with noise-like timing, and negative values indicate wrong-signed timing. Timing alpha measures entry and exit quality, not cumulative portfolio return. For the complementary schedule $q_{idt} = 1 - p_{idt}$, $q_{idt} - \bar{q}_{id} = -(p_{idt} - \bar{p}_{id})$, so $A_{id}(q, r) = -A_{id}(p, r)$. We use this identity only to diagnose directional structure; it does not establish a frictionless or executable counter-strategy.

To measure a distinct stock-selection margin, contemporaneous cross-sectional information coefficient (IC) correlates positions and returns across the set S_{dt} of stocks available to an

arm at the same decision time:

$$\mathrm{IC}_{dt} = \mathrm{Corr}_{i \in S_{dt}}(p_{idt}, r_{idt}). \tag{2}$$

Timing alpha therefore asks *when* the agent holds a given stock, whereas IC asks *which* stocks it prefers at a common timestamp. We keep the two estimands separate rather than interpreting one as a standardized version of the other.

3.2 Interventions on agent state

Throughout, *conditioning* means a change in representation or endogenous state constructed from the same permitted market information; it does not supply a new market variable. We vary representation, consequential state, narrative state, and latent state while preserving the market information set and binary action space. Writing $z_{idt}^{(c)}$ for the condition-specific state, each condition induces a policy $p_{idt} \sim \pi_c(\cdot \mid x_{id,1:t}, z_{idt}^{(c)})$. These parallel interventions let us ask which forms of endogenous state alter the structure and persistence of the decision trajectory.

Observation representation and multimodal reasoning. The numerical-text condition serializes the recent price path as explicit values, making local differences and ordering directly available to linguistic reasoning. The image condition renders the same underlying observations as a chart, making slope, curvature, extrema, and visual grouping jointly salient. It therefore tests whether a visual abstraction induces a different policy even when it contains no additional market variable. The multimodal condition presents both views simultaneously. This condition is not merely a redundant input check: it asks whether the agent reconciles numerical precision with chart-level geometry, privileges one representation when they suggest different heuristics, or constructs a decision that neither representation elicits alone. Differences among the three conditions therefore identify sensitivity to representation and multimodal integration while holding the underlying market variables fixed.

Consequential state. Stateful prompts progressively reveal the agent's own interaction with the price path. Current position and entry price supply immediate action state; a trade ledger supplies episodic memory of prior commitments; and an account summary adds outcome-laden feedback through realized and unrealized performance. This progression tests whether the same price evidence is interpreted differently once it is attached to the agent's own decisions and gains or losses. Because these variables are consequences of earlier actions, they can create a recursive policy in which a decision changes the state that conditions the next decision.

Endogenous narrative and latent state. Narrative self-conditioning makes the model's own prior interpretation part of its subsequent context without adding market information or updating parameters. Learned soft tokens provide a continuous latent state derived from the same observed price path. The following subsection develops these two conditioning channels as explicit and implicit counterparts.

3.3 Narrative self-conditioning and latent conditioning

Narrative self-conditioning exposes self-generated state in natural language. At each decision time, the model first writes a short note addressed to its future self. A second pass receives the current observation, the new note, and up to w earlier notes before emitting the action. Varying the visible history and instructions emphasizing conviction or consistency tests whether a recent commitment induces persistence rather than merely providing additional text.

The learned intervention supplies a parallel continuous, price-dependent state through soft-token embeddings:

$$x_{id,1:t} \xrightarrow{f_\theta} h_{idt} \xrightarrow{g_\phi} (e_1, \ldots, e_k) \longrightarrow \text{frozen VLM} \longrightarrow p_{idt}. \tag{3}$$

Here f_θ is a self-supervised price encoder and g_ϕ maps its state to a short continuous prefix. The encoder is trained to describe contemporaneous market state rather than forecast the

outcome: its representation captures trend, drawdown, and volatility but is not selected for forward-return prediction. Comparing narrative and latent conditioning asks whether persistence arises specifically from linguistic self-commitment or from giving the same frozen model an additional state representation more generally.

The GRU encoder is pretrained without decision labels using a VICReg-style objective that discourages collapsed representations (Bardes et al., 2022). Diagnostic heads verify that its state captures contemporaneous trend, drawdown, and volatility while failing to predict forward return. In the principal constrained adapter, each of two soft tokens is a convex combination of 92 frozen market-related token embeddings. Ablations replace these with unconstrained vectors, expand the basis to 492 embeddings, increase bandwidth to eight or sixteen soft tokens, or freeze the encoder and train only the 54k-parameter adapter. The soft tokens are inserted adjacent to the answer location, where a placement probe confirms that they affect the binary logit. Training uses a straight-through hard-forward/soft-backward objective; gradients pass through the frozen vision-language model, and a calibrated decision bias prevents all-long or all-flat collapse. Figure 2 summarizes the modality-specific encoding paths and the optional conditioning channels.

4 Experiments

4.1 Experimental protocol

Market panels and decision horizons. The source data provide each anonymized CSI-500 stock-day as a one-minute intraday price trace. Visual conditions receive the rolling one-minute chart available at the decision time; numerical conditions receive an indexed price serialization of the same history. The primary panel samples 23 decisions at 10-minute intervals per stock-day, and 5- and 20-minute panels provide horizon checks. The outcome is the subsequent non-overlapping return label for the selected interval; it is not an executable quote or a transaction-cost-adjusted return. The controlled memo and encoder evaluations use 1,500 stock-days, while larger trajectory sets support modality, state, and horizon checks. Every condition uses the same binary action semantics.

Prompt and state conditions. The principal conditions use Qwen3.5-9B, with Haiku and Granite providing model-family checks. Price paths are presented as numerical text, one-minute charts, or both. Stateful variants add the current position and entry price, a recent trade ledger, or an account summary. Memo actions use a two-pass protocol: the first pass writes the note, and the second receives that note plus up to $w \in \{0, 1, 5, 12\}$ earlier notes. The directive is either absent, emphasizes conviction, or asks for consistency with the earlier view unless evidence has clearly changed. Thus $w = 0$ retains the two-pass protocol and conditions the decision on the memo written at the current step, but supplies no earlier memo history; it is not a no-memo condition. All prompts retain identical long/flat action semantics. The standard agents and all narrative self-conditioning conditions, including the no-directive condition denoted "none," are inference-only interventions and require no training or parameter updates; their memos are generated online at evaluation time. Only the learned encoder–adapter conditions require an optimization stage.

Generation and parsing. The principal Qwen conditions use the locally cached `Qwen/Qwen3.5-9B` checkpoint in bfloat16. Standard agents are run through vLLM using the model's Hugging Face chat template with thinking disabled, temperature 0.7, top-p 0.95, and a 160-token output limit. Memo generation uses temperature 0.7, top-p 0.95, and a 220-token limit; its second-pass binary decision is greedy with a 24-token limit. Generation is batched across stock-days at each decision time while decisions within each stock-day remain sequential. The parser takes the final explicit `ACTION: 0/1` match and uses directional keywords only as a fallback. Standard-arm parse failures map to flat; failed memo decisions are recorded and excluded from scoring rather than treated as inherited holds. The stored experiment records preserve the checkpoint name and prompt implementation but not an immutable revision manifest.

Encoder optimization. The price encoder is a two-layer GRU with 256 hidden units, a 256-dimensional output, dropout 0.1, and a 120-minute lookback. Its seven channels are one-minute return, 10- and 30-minute volatility, 5- and 20-minute momentum, drawdown, and time of day. VICReg pretraining uses invariance and variance weights of 25 and covariance weight 1, with 10% magnitude jitter and 10% channel dropout. The principal adapter uses two soft tokens, a rank-64 bottleneck, and a 92-token lexical basis. AdamW uses learning rates 3×10^{-4} for the memo-conditioned adapter and 10^{-5} for the encoder, with weight decay 10^{-4}; an online scalar bias targets exposure 0.40.

Training and evaluation split. Training dates occupy the first three chronological fifths of the relevant panel. The memo-conditioned encoder checkpoint is evaluated without parameter updates on 1,500 stock-days from 28 later dates. The memo sweep uses 1,500 stock-days over 150 dates. All prompts and saved trajectories preserve chronological information availability.

Statistical inference and controls. The primary null draws random binary schedules at the same exposure. Unlike a price-level or momentum comparator, its measured timing alpha is near zero. Confidence intervals use a date-clustered percentile bootstrap: dates, rather than individual stock-days or decision times, are resampled. Cross-sectional IC is averaged over valid decision times; constant-position cross-sections are undefined and omitted. For an inverted position $q = 1 - p$ and comparator s, we also project the within-stock centered q on centered s and score the residual. This is a test of linear signal overlap, not causal independence or market optimality.

4.2 Do LLM agents exhibit systematic wrong-signed timing?

Figure 3 reports the complete set of standard conditions and two representative self-conditioning interventions rather than a selected modality subset. All 14 estimates and both displayed self-conditioning estimates are negative. The result spans three decision horizons, four state conditions, text and visual inputs, and a smaller Haiku cross-model check. Thus the central finding is not merely low or noisy performance: across the tested settings, the induced policies repeatedly exhibit wrong-signed within-stock timing. Text and chart modalities are not substitutes: the chart-only estimate is less negative, while the multimodal estimate is more negative than either. This is compatible with representation-dependent behavior, though it does not identify a psychological mechanism.

The state ablations do not exhibit a monotonic ordering. Adding position state makes the 5-minute effect more negative, ledger state weakens it, and full account state strengthens it again. At 20 minutes the same ordering changes. Self-relevant context therefore alters the behavioral policy, but "more context" is not a scalar treatment. Because these runs differ in horizon, sample, and model, we treat the repeated sign as robustness evidence and do not pool their magnitudes.

The estimated contrast depends on the benchmark. On the principal text sample, matching exposure to a random schedule gives an LLM-minus-null effect of -46.2 bps (95% CI $[-50.5, -41.8]$). Replacing that null with a price-level or last-return rule changes the comparison dramatically because those schedules have nonzero timing of their own. We consequently use random exposure matching for the behavioral claim and reserve technical signals for overlap analysis.

4.3 Is the wrong-signed schedule trivial?

The complementary schedule makes the sign of the behavior explicit but is algebraically determined by Equation 1. The more informative question is whether the tested conventional price signals linearly account for the induced schedule. We invert the text agent and remove the within-stock linear component explained by each comparator. Table 1 shows substantial residual positive timing after projection on a GRU price model, a 38-feature gradient-boosted tree, and a one-period reversal rule. Thus, within this linear-overlap diagnostic, the tested

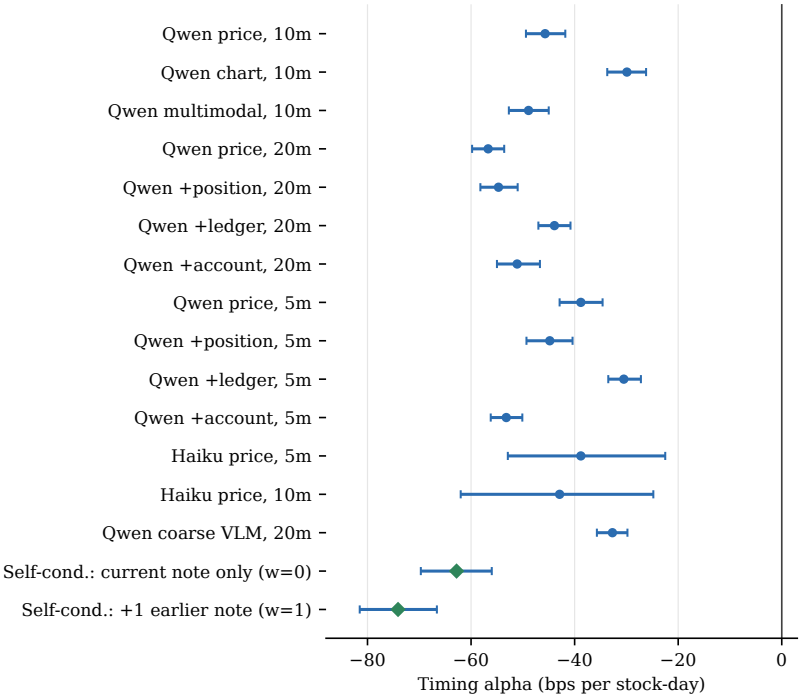


Figure 3: Unimodal and multimodal input conditions together with two representative narrative self-conditioning interventions. Diamonds identify self-conditioning conditions. Points report within-stock timing alpha with 95% date-clustered confidence intervals. Sample sizes and exact values are reported in Appendix A.

Table 1: Inverted text-agent timing after within-stock projection on conventional price signals.

Comparator	Stock-days	Corr.	Inverted	Residual timing (95% CI)
GRU price model	2,519	+0.114	+46.8	+43.9 [38.1, 49.5]
GBDT-38	13,717	+0.132	+45.7	+40.4 [37.0, 43.8]
One-period reversal	13,717	+0.286	+45.7	+38.9 [36.0, 41.8]

signals do not fully account for the structured wrong-signed schedule. This establishes that the tested linear projections leave directional structure in the complementary schedule.

4.4 How does self-conditioning alter structured failure?

Figure 4 shows the memo-window sweep. The largest negative estimate occurs with exactly one prior note. A condition without visible history provides no earlier commitment to defend, while a longer visible history may make inconsistency easier to recognize. The one-earlier-note condition also has the fewest position changes and the highest rate of explicit self-reference in generated memos. These observations are consistent with self-conditioning reinforcing an existing interpretation and increasing policy persistence; they are mechanistic correlates, not proof that persistence mediates the full timing change.

Figure 4 reports the temporal result together with a cross-sectional diagnostic. In particular, timing asks when an agent holds a stock, whereas IC asks which stocks it prefers at the same timestamp.

Adding the chart modality does not significantly improve narrative self-conditioning. Moreover, rendering a frozen encoder's highest-weight vocabulary items as literal tokens is worse

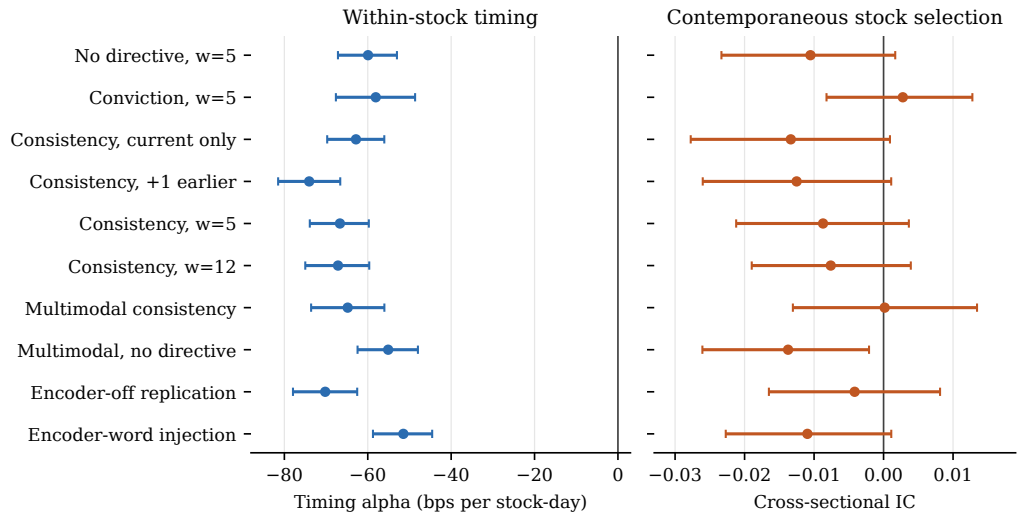


Figure 4: Within-stock timing alpha and contemporaneous cross-sectional IC for ten narrative self-conditioning conditions, each evaluated on 1,500 stock-days. Both panels report 95% date-clustered confidence intervals. Self-conditioning conditions substantially reorder timing performance while IC remains near zero. Exact values and later-period estimates are reported in Appendix A.

than a matched encoder-off memo control (-51.5 versus -70.2). The result indicates that the adapter's embedding-space representation is not preserved by a direct lexical rendering.

Appendix B compares this narrative channel with the supporting latent-steering ablations; the saved configurations show that added latent capacity does not supply a monotonic account of the effect.

Timing versus cross-sectional selection. The self-conditioning conditions make the distinction between timing and IC visible. Their timing estimates differ substantially, while IC remains close to zero: the $w = 1$ condition has IC -0.0125 with CI $[-0.0260, 0.0011]$. The only memo IC among the ten conditions whose interval excludes zero is the image/no-directive condition, which does not have the largest timing effect. Thus the tested self-conditioning interventions reorganize when a given stock is held more strongly than which stock is preferred at a timestamp. Supplementary correlations across experimental arms appear in Appendix C; they summarize metric coupling rather than probabilistic independence.

5 Conclusion

RetailAgent exposes a form of structured sequential failure: across the tested standard conditions, its long/flat decision trajectories have negative within-stock timing rather than merely near-zero timing. The pattern spans modality, horizon, endogenous account state, and model checks. Narrative self-conditioning coincides with greater persistence and stronger wrong-signed timing.

The scientific contribution is not a promise of retail imitation, market exploitation, or executable alpha. It is a reproducible demonstration of directional behavioral error, diagnosed against an exposure-matched null and the linear components of tested conventional signals. Because directional error can define a response against which another decision maker could optimize, it motivates a future leader–follower or bilevel-inspired counter-policy problem, which we do not solve. Confirmation requires an untouched date range, a second prespecified model family, executable returns, and common cost-aware portfolio construction.

Acknowledgments

Anonymous for review.

AI use statement

Large language models are themselves the experimental subjects in this work, as described in the methodology; this use is part of the research design rather than author assistance. Separately, the authors used generative AI tools as research assistants for coding and refining the experiment protocol, assisting with literature discovery, and drafting and editing portions of the manuscript. All AI-assisted research suggestions and manuscript content were reviewed by the authors, who take responsibility for the final claims and content.

Reproducibility statement

The accompanying artifact provides the implementation of the experimental protocol, prompts, panel construction, scoring, and statistical evaluation. The primary metric is defined in Equation 1, and all reported uncertainty uses date-clustered bootstrap resampling. The paper documents the model configurations and experimental procedures needed to reproduce the analysis given equivalent input data.

References

Gati V. Aher, Rosa I. Arriaga, and Adam Tauman Kalai. Using large language models to simulate multiple humans and replicate human subject studies. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 337–371. PMLR, 2023. URL https://proceedings.mlr.press/v202/aher23a.html.

Lisa P Argyle, Ethan C Busby, Nancy Fulda, Joshua R Gubler, Christopher Rytting, and David Wingate. Out of one, many: Using language models to simulate human samples. *Political Analysis*, 31(3):337–351, 2023.

Brad M Barber and Terrance Odean. Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The journal of Finance*, 55(2): 773–806, 2000.

Brad M Barber and Terrance Odean. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies*, 21(2):785–818, 2008.

Brad M. Barber and Terrance Odean. The behavior of individual investors. In *Handbook of the Economics of Finance*, volume 2, pp. 1533–1570. Elsevier, 2013. doi: 10.1016/B978-0-44-459406-8.00022-6.

Adrien Bardes, Jean Ponce, and Yann LeCun. VICReg: Variance-invariance-covariance regularization for self-supervised learning. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=xm6YD62D1Ub.

Hedi Benamar. To see is to know: Simultaneous display of market data for retail investors. *Review of Finance*, 23(2):397–437, 2019. doi: 10.1093/rof/rfy007.

Sanford J Grossman and Joseph E Stiglitz. On the impossibility of informationally efficient markets. *The American economic review*, 70(3):393–408, 1980.

John J Horton, Apostolos Filippas, and Benjamin S Manning. Large language models as simulated economic agents: What can we learn from homo silicus? Technical report, National Bureau of Economic Research, 2023.

Tiancheng Hu and Nigel Collier. Quantifying the persona effect in LLM simulations. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 10289–10307, 2024. doi: 10.18653/v1/2024.acl-long.554.

Tiancheng Hu, Joachim Baumann, Lorenzo Lupo, Nigel Collier, Dirk Hovy, and Paul Röttger. SIMBENCH: Benchmarking the ability of large language models to simulate human behaviors. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/2fc4d6a6400addda7493b9db258c34e0-Paper-Conference.pdf.

Soeren Hvidkjaer. Small trades and the cross-section of stock returns. *The Review of Financial Studies*, 21(3):1123–1151, 2008.

Michael Krumdick, Rik Koncel-Kedziorski, Viet Dac Lai, Varshini Reddy, Charles Lovering, and Chris Tanner. BizBench: A quantitative reasoning benchmark for business and finance. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 8309–8332, 2024. doi: 10.18653/v1/2024.acl-long.452.

Albert S Kyle. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pp. 1315–1335, 1985.

Brian Lester, Rami Al-Rfou, and Noah Constant. The power of scale for parameter-efficient prompt tuning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 3045–3059, 2021. doi: 10.18653/v1/2021.emnlp-main.243.

Xiang Lisa Li and Percy Liang. Prefix-tuning: Optimizing continuous prompts for generation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pp. 4582–4597, 2021. doi: 10.18653/v1/2021.acl-long.353.

Yuan Li, Bingqiao Luo, Qian Wang, Nuo Chen, Xu Liu, and Bingsheng He. Cryptotrade: A reflective llm-based agent to guide zero-shot cryptocurrency trading. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pp. 1094–1106, 2024.

Ryan Liu, Jiayi Geng, Joshua Peterson, Ilia Sucholutsky, and Thomas L. Griffiths. Large language models assume people are more rational than we really are. In *International Conference on Learning Representations*, 2025a. URL https://mlanthology.org/iclr/2025/liu2025iclr-large/.

Shu Liu, Shangqing Zhao, Chenghao Jia, Xinlin Zhuang, Zhaoguang Long, Jie Zhou, Aimin Zhou, Man Lan, and Yang Chong. FinDABench: Benchmarking financial data analysis ability of large language models. In *Proceedings of the 31st International Conference on Computational Linguistics*, pp. 710–725, 2025b. URL https://aclanthology.org/2025.coling-main.48/.

Xiao Liu, Hao Yu, Hanchen Zhang, Yifan Xu, Xuanyu Lei, Hanyu Lai, Yu Gu, Hangliang Ding, Kaiwen Men, Kejuan Yang, Shudan Zhang, Xiang Deng, Aohan Zeng, Zhengxiao Du, Chenhui Zhang, Sheng Shen, Tianjun Zhang, Yu Su, Huan Sun, Minlie Huang, Yuxiao Dong, and Jie Tang. AgentBench: Evaluating LLMs as agents. In *International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=zAdUB0aCTQ.

Xiao-Yang Liu, Guoxuan Wang, Hongyang Yang, and Daochen Zha. Fingpt: Democratizing internet-scale data for financial large language models. *arXiv preprint arXiv:2307.10485*, 2023.

Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, pp. 1–22. ACM, 2023. doi: 10.1145/3586183.3606763.

Xueqing Peng, Lingfei Qian, Yan Wang, Ruoyu Xiang, Yueru He, Yang Ren, Mingyang Jiang, Vincent Jim Zhang, Yuqing Guo, Jeff Zhao, Huan He, Yi Han, Yun Feng, Yuechen Jiang, Yupeng Cao, Haohang Li, Yangyang Yu, Xiaoyu Wang, Penglei Gao, Shengyuan Lin, Keyi Wang, Shanshan Yang, Yilun Zhao, Zhiwei Liu, Peng Lu, Jerry Huang, Suyuchen Wang, Triantafillos Papadopoulos, Polydoros Giannouris, Efstathia Soufleri, Nuo Chen, Zhiyang Deng, Heming Fu, Yijia Zhao, Mingquan Lin, Meikang Qiu, Kaleb E. Smith, Arman Cohan, Xiao-Yang Liu, Jimin Huang, Guojun Xiong, Alejandro Lopez-Lira, Xi Chen, Junichi Tsujii, Jian-Yun Nie, Sophia Ananiadou, and Qianqian Xie. MultiFinBen: Benchmarking large language models for multilingual and multimodal financial application. In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 16934–16963, 2026. doi: 10.18653/v1/2026.acl-long.770. URL https://aclanthology.org/2026.acl-long.770/.

Giorgio Piatti, Zhijing Jin, Max Kleiman-Weiner, Bernhard Schölkopf, Mrinmaya Sachan, and Rada Mihalcea. Cooperate or collapse: Emergence of sustainable cooperation in a society of LLM agents. In *Advances in Neural Information Processing Systems*, volume 37, pp. 111715–111759, 2024. doi: 10.52202/079017-3548. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/ca9567d8ef6b2ea2da0d7eed57b933ee-Abstract-Conference.html.

Meir Statman, Steven Thorley, and Keith Vorkink. Investor overconfidence and trading volume. *The Review of Financial Studies*, 19(4):1531–1565, 2006. doi: 10.1093/rfs/hhj032.

Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. Bloomberggpt: A large language model for finance. *arXiv preprint arXiv:2303.17564*, 2023.

Yijia Xiao, Edward Sun, Di Luo, and Wei Wang. Tradingagents: Multi-agents llm financial trading framework. *arXiv preprint arXiv:2412.20138*, 2024.

Chengxing Xie, Canyu Chen, Feiran Jia, Ziyu Ye, Shiyang Lai, Kai Shu, Jindong Gu, Adel Bibi, Ziniu Hu, David Jurgens, James Evans, Philip H. S. Torr, Bernard Ghanem, and Guohao Li. Can large language model agents simulate human trust behavior? In *Advances in Neural Information Processing Systems*, volume 37, pp. 15674–15729, 2024a. doi: 10.52202/079017-0501. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/1cb57fcf7ff3f6d37eebae5becc9ea6d-Abstract-Conference.html.

Qianqian Xie, Weiguang Han, Zhengyu Chen, Ruoyu Xiang, Xiao Zhang, Yueru He, Mengxi Xiao, Dong Li, Yongfu Dai, Duanyu Feng, Yijing Xu, Haoqiang Kang, Ziyan Kuang, Chenhan Yuan, Kailai Yang, Zheheng Luo, Tianlin Zhang, Zhiwei Liu, Guojun Xiong, Zhiyang Deng, Yuechen Jiang, Zhiyuan Yao, Haohang Li, Yangyang Yu, Gang Hu, Jiajia Huang, Xiao-Yang Liu, Alejandro Lopez-Lira, Benyou Wang, Yanzhao Lai, Hao Wang, Min Peng, Sophia Ananiadou, and Jimin Huang. FinBen: A holistic financial benchmark for large language models. In *Advances in Neural Information Processing Systems*, volume 37, pp. 95716–95743, 2024b. doi: 10.52202/079017-3033. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/adb1d9fa8be4576d28703b396b82ba1b-Abstract-Datasets_and_Benchmarks_Track.html.

Hao Yang, Junyang Lin, An Yang, Peng Wang, and Chang Zhou. Prompt tuning for unified multimodal pretrained models. In *Findings of the Association for Computational Linguistics: ACL 2023*, pp. 402–416, 2023. doi: 10.18653/v1/2023.findings-acl.27.

Yangyang Yu, Haohang Li, Zhi Chen, Yuechen Jiang, Yang Li, Denghui Zhang, Rong Liu, Jordan W. Suchow, and Khaldoun Khashanah. FinMem: A performance-enhanced LLM trading agent with layered memory and character design. *Proceedings of the AAAI Symposium Series*, 3(1):595–597, 2024a. doi: 10.1609/aaaiss.v3i1.31290.

Yangyang Yu, Zhiyuan Yao, Haohang Li, Zhiyang Deng, Yuechen Jiang, Yupeng Cao, Zhi Chen, Jordan W. Suchow, Zhenyu Cui, Rong Liu, Zhaozhuo Xu, Denghui Zhang, Koduvayur Subbalakshmi, Guojun Xiong, Yueru He, Jimin Huang, Dong Li, and Qianqian Xie. FinCon: A synthesized LLM multi-agent system with conceptual

verbal reinforcement for enhanced financial decision making. In *Advances in Neural Information Processing Systems*, volume 37, pp. 137010–137045, 2024b. doi: 10.52202/079017-4354. URL https://proceedings.neurips.cc/paper/2024/hash/f7ae4fe91d96f50abc2211f09b6a7e49-Abstract-Conference.html.

A Numerical results underlying the figures

Table A2: Exact values underlying Figure 3.

Model/observation	Interval	Stock-days	Timing alpha (95% CI)
Qwen price	10m	13,710	−45.7 [−49.4, −41.8]
Qwen chart	10m	14,937	−29.9 [−33.7, −26.2]
Qwen multimodal (text+chart)	10m	14,438	−48.9 [−52.7, −45.0]
Qwen price	20m	23,594	−56.7 [−59.8, −53.6]
Qwen +position	20m	25,410	−54.7 [−58.2, −51.0]
Qwen +ledger	20m	25,718	−43.9 [−47.0, −40.8]
Qwen +account	20m	22,013	−51.1 [−55.0, −46.7]
Qwen price	5m	9,937	−38.8 [−42.9, −34.6]
Qwen +position	5m	9,961	−44.8 [−49.3, −40.4]
Qwen +ledger	5m	9,966	−30.5 [−33.5, −27.2]
Qwen +account	5m	9,052	−53.2 [−56.2, −50.1]
Haiku price	5m	827	−38.8 [−52.9, −22.5]
Haiku price	10m	165	−42.9 [−62.0, −24.8]
Qwen coarse-image VLM	20m	28,387	−32.7 [−35.7, −29.8]
Self-conditioning: current note only ($w = 0$)	10m	1,500	−62.8 [−69.7, −56.0]
Self-conditioning: +1 earlier note ($w = 1$)	10m	1,500	−74.1 [−81.5, −66.6]

Table A3: Remaining memo timing results underlying Figure 4; the $w = 0$ and $w = 1$ consistency conditions are included in the preceding consolidated table. Each condition uses 1,500 stock-days.

Self-conditioning condition	Timing alpha (95% CI)	Cross-sectional IC (95% CI)
No directive, $w = 5$	−59.9 [−67.1, −53.0]	−0.0105 [−0.0233, 0.0017]
Conviction, $w = 5$	−58.1 [−67.6, −48.6]	+0.0028 [−0.0082, 0.0128]
Consistency, $w = 5$	−66.7 [−73.9, −59.7]	−0.0087 [−0.0212, 0.0036]
Consistency, $w = 12$	−67.1 [−75.0, −59.6]	−0.0076 [−0.0190, 0.0039]
Multimodal consistency	−64.8 [−73.6, −56.0]	+0.0002 [−0.0130, 0.0134]
Multimodal, no directive	−55.1 [−62.4, −48.0]	−0.0137 [−0.0261, −0.0021]
Encoder-off replication	−70.2 [−77.9, −62.5]	−0.0042 [−0.0165, 0.0081]
Encoder-word injection	−51.5 [−58.8, −44.5]	−0.0110 [−0.0227, 0.0011]

Table A4: Later-period and persistence diagnostics for the memo-window sweep.

Self-conditioning condition	Main-period alpha (95% CI)	Later period	Changes/stock-day
No directive, $w = 5$	−59.9 [−67.4, −52.9]	−65.3	2.5
Consistency, current only ($w = 0$)	−62.8 [−69.7, −56.1]	−68.3	2.5
Consistency, +1 earlier ($w = 1$)	−74.1 [−81.5, −66.8]	−85.6	1.7
Consistency, $w = 5$	−66.7 [−74.2, −59.8]	−82.9	1.8
Consistency, $w = 12$	−67.1 [−74.9, −59.3]	−79.3	1.9

B Latent-steering evaluations

Table A6 retains the objective used by each saved configuration. Passive, I, K8, K16, and Vocab-500 use the cross-sectionally centered outcome; F, G, H, and J use alternative transformed objectives. Therefore the table documents the saved ablations but does not provide a single numerically comparable capacity ranking.

Table A5: Exact condition-level correlations underlying Figure A5.

Metric pair	Conventional conditions Pearson	Conventional conditions Spearman	Self-conditioning conditions Pearson	Self-conditioning conditions Spearman
Timing vs portfolio return	+0.355	+0.392	−0.066	+0.018
Timing vs cross-sectional IC	+0.684	+0.399	−0.075	−0.212
Portfolio return vs cross-sectional IC	+0.859	+0.881	+0.920	+0.867

Table A6: Latent-steering evaluations. Reported scores retain the objective used by each saved configuration.

Arm	Configuration	Reported score (95% CI)
Passive	same prompt, no soft tokens	−21.0 [−23.8, −18.3]
F	92-word basis, trainable encoder	−38.4 [−43.6, −33.3]
G	frozen encoder, 54k adapter	−39.7 [−46.8, −32.4]
H	unconstrained free vectors	−2.1 [−6.5, +2.6]
I	$k = 2$, 92-word basis	−33.7 [−37.5, −30.0]
J	combined transformation	−28.7 [−32.3, −25.0]
K8	eight soft tokens	−32.3
K16	sixteen soft tokens	−29.2
Vocab-500	492-word basis	−33.8 [−37.5, −30.1]

The passive prompt scores −21.0, while the $k = 2$ steered arm scores −33.7, so the learned channel can change decisions. Among the directly comparable configurations, increasing token count or lexical basis size does not yield a monotonic pattern, whereas fully unconstrained vectors score near zero. The memo-conditioned checkpoint, trained inside the two-pass memo procedure, scores −15.4 (95% CI [−24.6, −5.1]) on 1,500 stock-days from 28 evaluation dates, with exposure 0.397 and 6.71 position changes per stock-day. Its different prompt, sample, and training procedure preclude a causal comparison with memo-only arms.

C Cross-arm metric coupling

For every experimental condition with complete trajectories, we compute timing alpha, contemporaneous cross-sectional portfolio return, and cross-sectional IC. Figure A5 reports correlations *across arms*. For 12 conventional conditions, timing alpha is associated with cross-sectional IC (Pearson 0.684), whereas the corresponding association is not present in the ten-condition memo sweep (Pearson −0.075 and Spearman −0.212). Cross-sectional portfolio return and IC remain strongly associated in both families (0.859 and 0.920). These correlations summarize metric coupling across the reported arms; they do not establish probabilistic independence or a distinct latent factor.

D Illustrative decision trajectories

The following cases are selected ex post to make the estimand and the memo mechanism concrete; they are illustrations rather than additional statistical evidence. The first A6 is a saved price-only text trajectory with especially transparent adverse timing. The agent was long in 7 of 23 intervals. Returns during its long intervals sum to −576.7 bps, whereas returns during flat intervals sum to +575.9 bps, producing timing alpha of −576.4 bps and a long/flat account change of −5.64%. That account change compounds the return label only while the agent is long.

Table A7 reports a second saved trajectory from the one-note memo condition. We report the generated memo and observable action sequence, not an unobserved chain of thought. The stock rose strongly over the day, so the long intervals had positive total return; nevertheless, the agent remained flat through much of the rise and entered before comparatively weaker intervals. This yields positive long-only return but timing alpha of −867.3 bps, directly

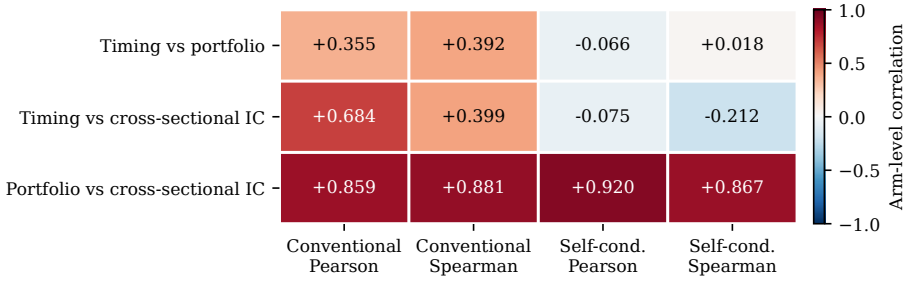


Figure A5: Correlations among condition-level metrics. "Conventional" contains 12 conditions without narrative self-conditioning; "self-conditioning" contains ten two-pass conditions evaluated on 1,500 stock-days each. The correlations summarize metric coupling across conditions rather than probabilistic independence of random variables. Exact values appear in the preceding table.

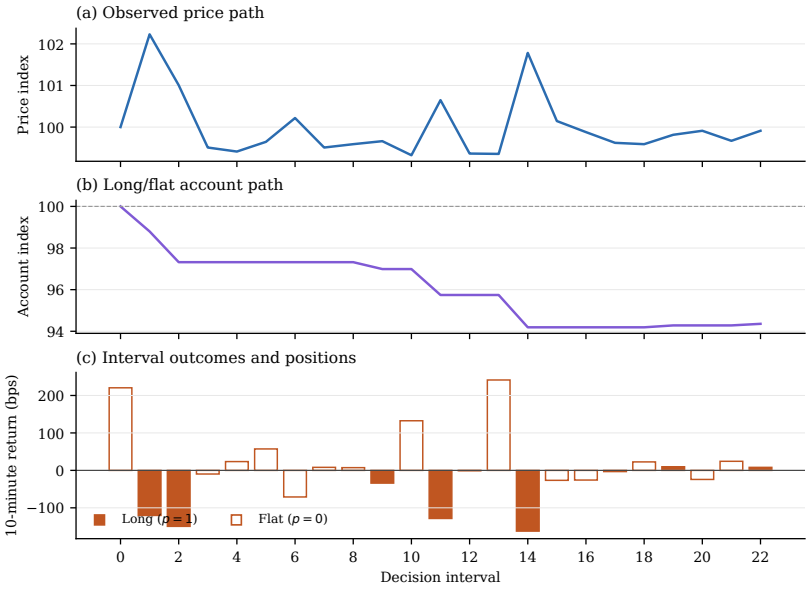


Figure A6: Example of a bad timing trajectory.

illustrating why a cross-sectional participation measure cannot replace the within-stock timing estimand.

E Cross-model sign checks

Table A8 records model-family checks separately from the main Qwen analysis. Haiku reproduces the wrong sign at two decision intervals despite small samples, and Granite shows the same direction across four state prompts.

Remark. The Granite point estimates were recovered from summary records after a cache migration; the corresponding confidence intervals were not recoverable.

F Trajectory-shuffling controls

We assess whether the negative timing estimate relies on the original temporal pairing of actions and returns by shuffling the saved actions from the principal 10-minute price-text

Table A7: Illustrative memo and observable decision trace (anonymized stock 303, date 249; consistency instruction, $w = 1$).

Field	Saved content or derived trajectory summary
Action sequence	Flat for decisions 1–7; switches to long at decision 8; remains long through decision 23 (one position change; exposure 0.696).
Memo excerpt	"My bullish thesis remains fully intact. The decisive breakout above 108.71 is confirmed, and the recent pullback to 114.39 served only as healthy consolidation rather than a reversal signal."
Persistence	The memo explicitly preserves the prior bullish interpretation, and the subsequent action sequence contains no exit despite multiple negative intervals.
Outcome summary	Long-interval return +359.4 bps; flat-interval return +1,404.0 bps; timing alpha −867.3 bps.

Table A8: Cross-model timing-alpha checks.

Model	Observation	Interval	Stock-days	Timing alpha
Qwen3.5-4B	price text	10m	12,784	−53.2
Qwen3.5-9B	price text	10m	13,710	−45.7
Claude Haiku 4.5	price text	5m	827	−38.8
Claude Haiku 4.5	price text	10m	165	−42.9
Phi-4-mini	price text	10m	14,993	−28.4
Granite-4-h-small	price text	10m	600	−36.3
Granite-4-h-small	+position	10m	600	−25.6
Granite-4-h-small	+ledger	10m	600	−14.1
Granite-4-h-small	+account	10m	600	−8.2

condition while holding the return paths fixed. The intact trajectory yields $\alpha = -45.7$ bps per stock-day, compared with −3.5 bps after global shuffling and −8.7 bps after same-day shuffling. Since these perturbations retain the realized action mix but alter where those actions occur along the price path, the sharp attenuation suggests that most of the timing magnitude comes from the original sequential action–return alignment rather than from long exposure alone.

G Static LoRA and inverted pool hedge

We additionally trained a static low-rank adaptation of Qwen3.5-4B as a deliberate stress test of the behavioral effect. The adapter has rank four on the query, key, value, and output projections in four transformer layers (393,216 trainable parameters). It was optimized on 6,000 stock-days to minimize the soft within-stock timing objective, using the probability obtained from the logits of the binary action tokens. This differs from the inference-only conditions in the main study: the LoRA condition was explicitly trained to amplify wrong-signed timing.

We evaluate the first-epoch adapter by direct binary-token readout on 12,000 stock-days: 200 stocks on each of 60 dates excluded from the adapter's training, validation, and test partitions. Its timing alpha is −72.90 bps per stock-day (95% date-bootstrap CI $[-77.70, -68.03]$), its soft timing alpha is −71.32 bps ($[-75.69, -66.85]$), and its mean within-cell temporal IC is −0.2213 ($[-0.2354, -0.2066]$). Hard exposure is 0.505 and the mean number of position changes is 2.34 per stock-day.

Table A9 evaluates the complementary position $1 - p_{idt}$ as a portfolio diagnostic. At each decision time, the strategy holds an equal-weight unit long portfolio of stocks selected by the inverted signal and shorts one unit of a common contemporaneous pool hedge. The selected-pool hedge is the equal-weight return of the 200 stocks assigned on that date; the full-pool hedge uses all 500 stocks. Transaction costs are charged in basis points per unit of one-way turnover. Long-book turnover and the single hedge instrument are charged

separately, so the hedge is charged once rather than once for every selected stock. Initial entry is included and end-of-day liquidation is not. Mean one-way turnover is 4.65 per day under both hedge definitions.

Table A9: Inverted LoRA direct-readout portfolio as a function of one-way transaction friction. Return is geometrically annualized from 60 daily observations; Sharpe is annualized by $\sqrt{252}$.

	Selected 200-stock pool		Full 500-stock pool	
Friction (bp)	Annual return	Sharpe	Annual return	Sharpe
0	+49.27%	+8.71	+50.24%	+8.56
1	+32.76%	+6.29	+33.63%	+6.24
2	+18.08%	+3.77	+18.85%	+3.80
3	+5.02%	+1.15	+5.71%	+1.26
4	−6.59%	−1.57	−5.98%	−1.38
5	−16.92%	−4.39	−16.38%	−4.14

The point estimate crosses zero between three and four basis points of one-way friction (approximately 3.4–3.5 bp). These values use return labels rather than executable prices and omit borrow, impact, latency, and hedge-basis risk. The inversion is also selected after establishing the sign of the trained objective. The table therefore quantifies the cost sensitivity of a saved behavioral signal; it is not evidence of a deployable return at the reported annualized magnitude.

H Threats to validity

Three limitations bound the interpretation. First, RetailAgent is a controlled behavioral proxy, not an observed retail population. Second, borrow, impact, latency, and mapping to executable prices lie outside the study. Third, the historical market does not react to RetailAgent: the experiment contains neither strategic counterparties nor endogenous order flow or price feedback.

I Additional reporting conventions

Alpha is cumulative within a stock-day and then averaged across stock-days; it is not standardized by exposure. The inverse schedule is used as a diagnostic of sign, not as evidence of frictionless short implementation. Unless explicitly stated, reported comparisons are descriptive across arms and causal language is reserved for matched prompts, samples, rewards, and inference procedures.