Too Smart to be Human: Can AI Agents Replace Us in Behavioral Experiments?

By: John Garcia
California Lutheran University

NSF ACCESS Regional AI Workshop

January 22, 2026

Paper Link (SSRN): https://bit.ly/497IyU3

California Lutheran
UNIVERSITY

The Perfect Laboratory for Investor Psychology?

The Problem

For decades, we've struggled to isolate variables in the messy world of human finance. Human subjects are noisy, expensive, and hard to control.

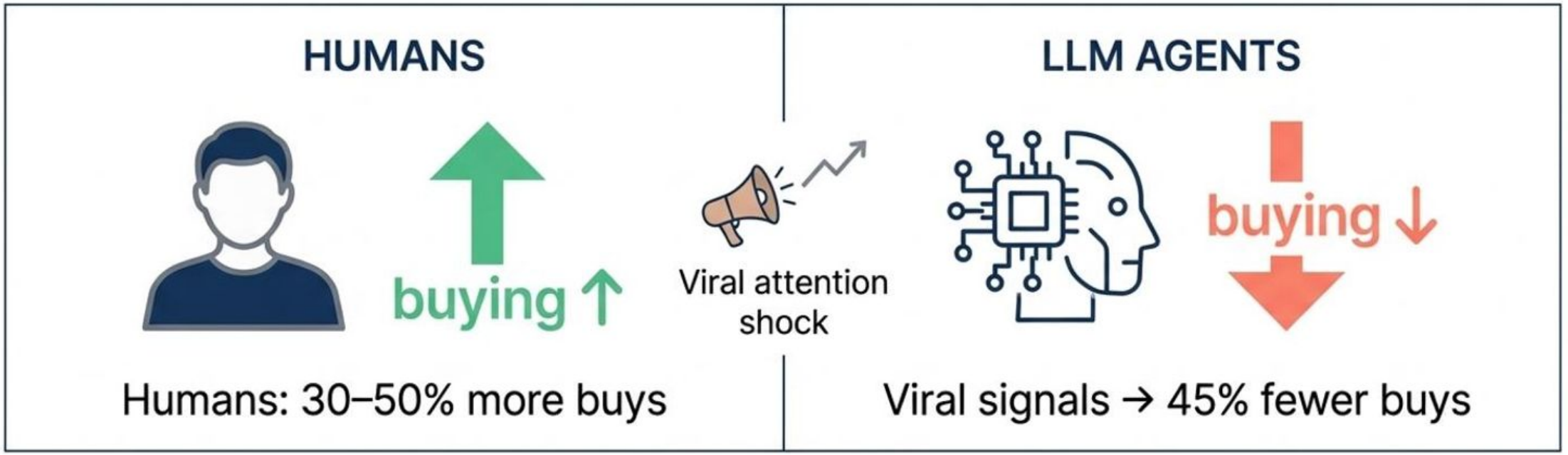
The Promise

Enter **Homo Silicus:** A new species of scalable, perfectly controllable synthetic investors. The promise is immense: finally, a clean environment to test causal drivers of market behavior.

The Test: Do they replicate Attention-Driven Trading (buying what's popular)?.

The Surprise: Viral Signals Kill Demand (-45%)

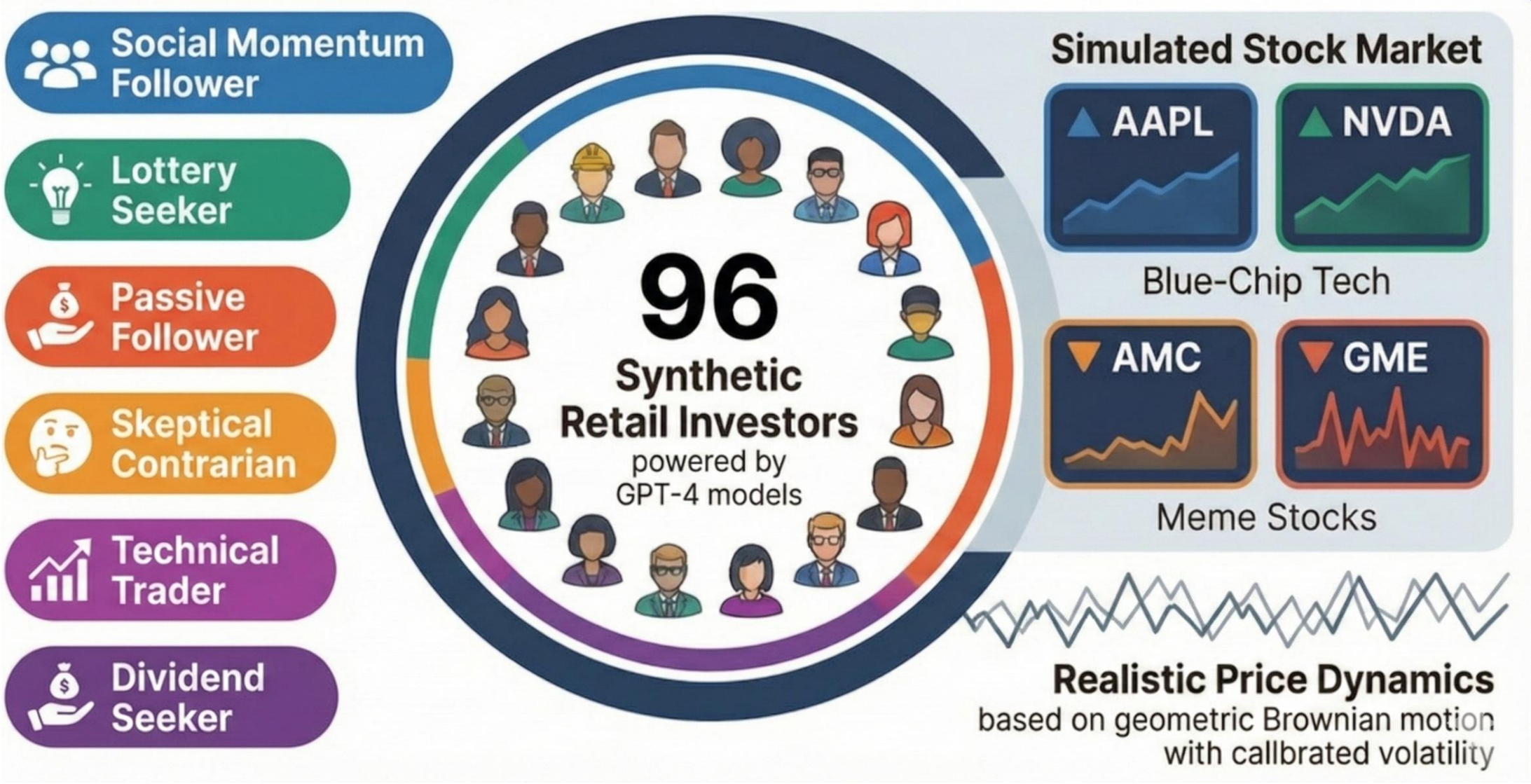
Behavioral finance research finds that when a stock suddenly becomes highly visible—news coverage, trending posts, "going viral"— investors tend to buy more. In this experiment, LLM agents did the opposite.



Same signal, opposite response

The Driver: Humans trade on FOMO (Emotion); LLMs trade on Skepticism (Rationality)

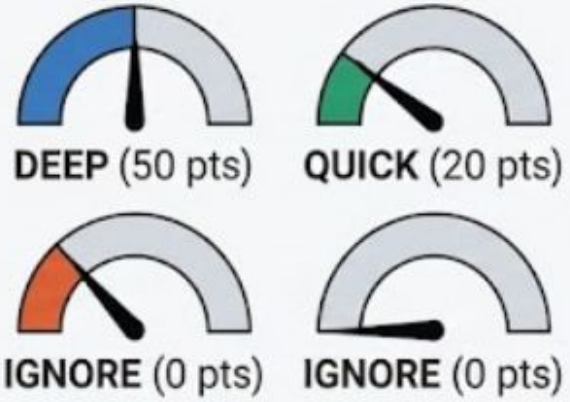
The Setup: Building the Agents & Market



The Mechanism: The 252-Day Trading Simulation

Stage 1: Allocate Scarce Attention

> 100-point budget

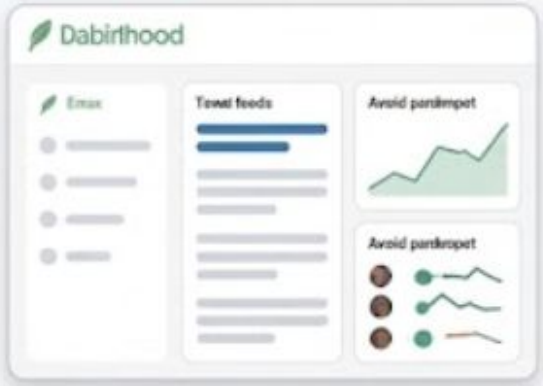


DEEP (50 pts) **QUICK** (20 pts)

IGNORE (0 pts) **IGNORE** (0 pts)

For each of four stocks

Stage 2: View Information



Personalized news feeds for researched stocks only.

Stage 3: Make a Trading Decision

Portfolio

BUY

SELL

HOLD

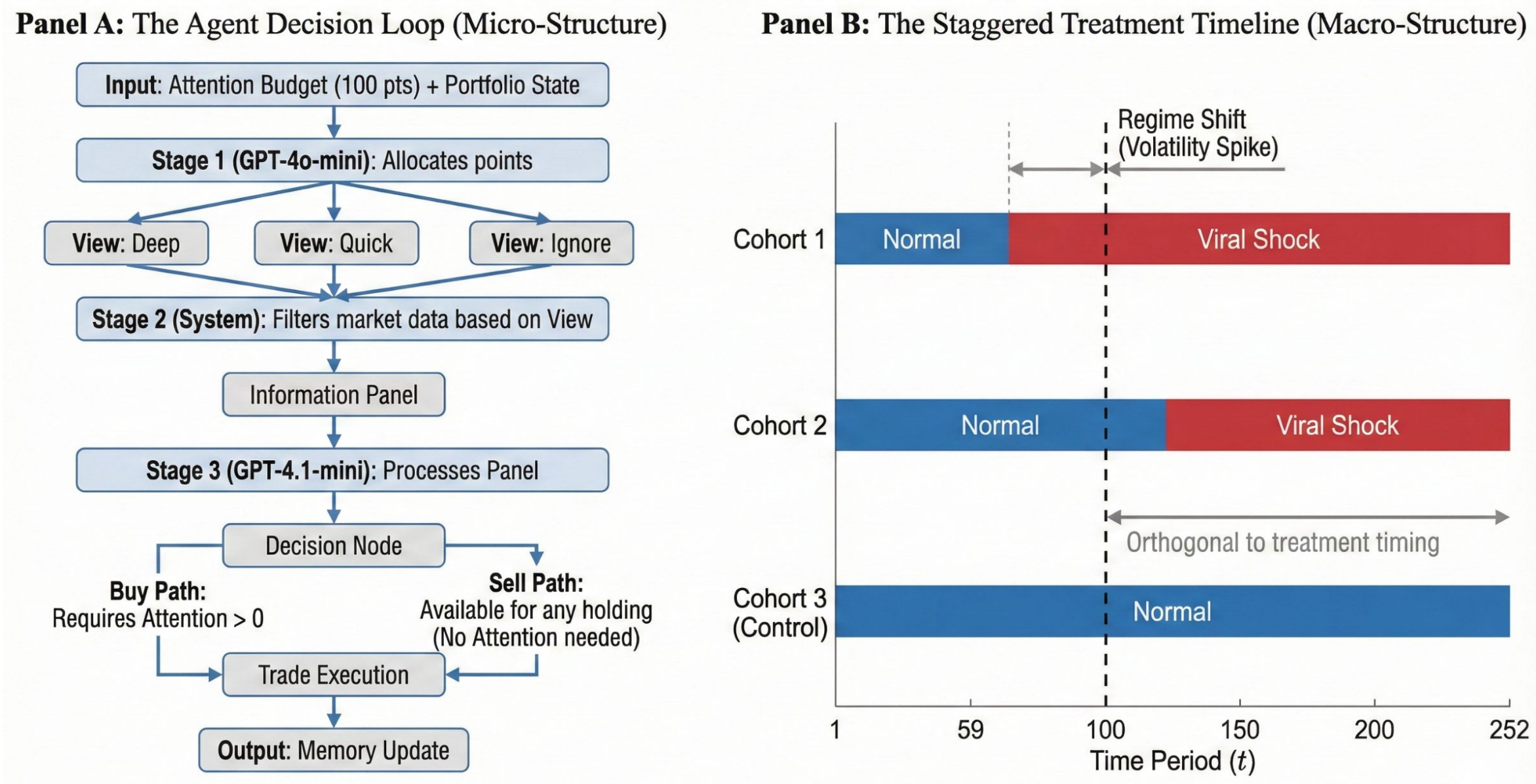
Personalized news in stocks on a portfolio

Asymmetric Trading Rule:

Agents **MUST** allocate attention to a stock to **BUY**.

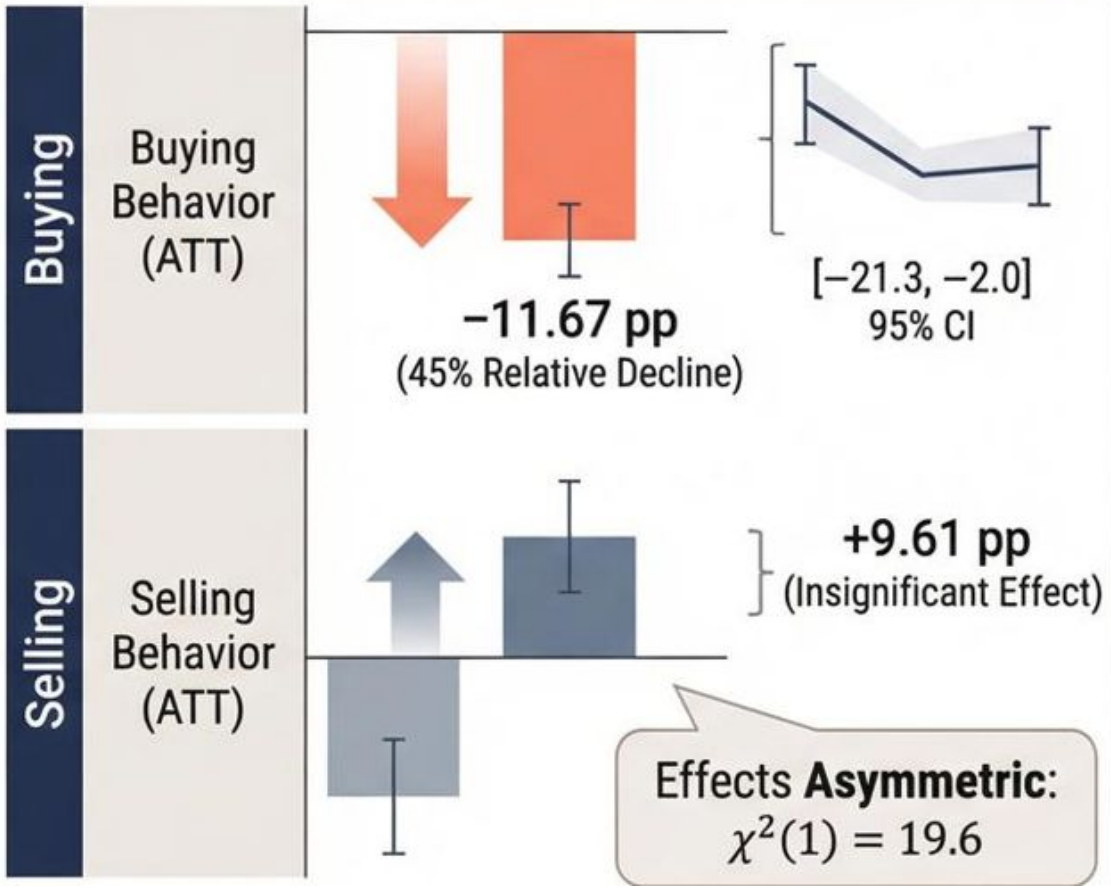
Can **SELL** without prior research.

Experimental Design: Staggered Attention



The Causal Effect: Buying Drops 45%

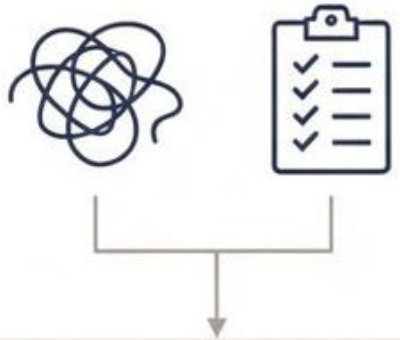
Primary Causal Estimates



Practical Impact & Prevention Rate

One purchase prevented per ~9 exposed agents

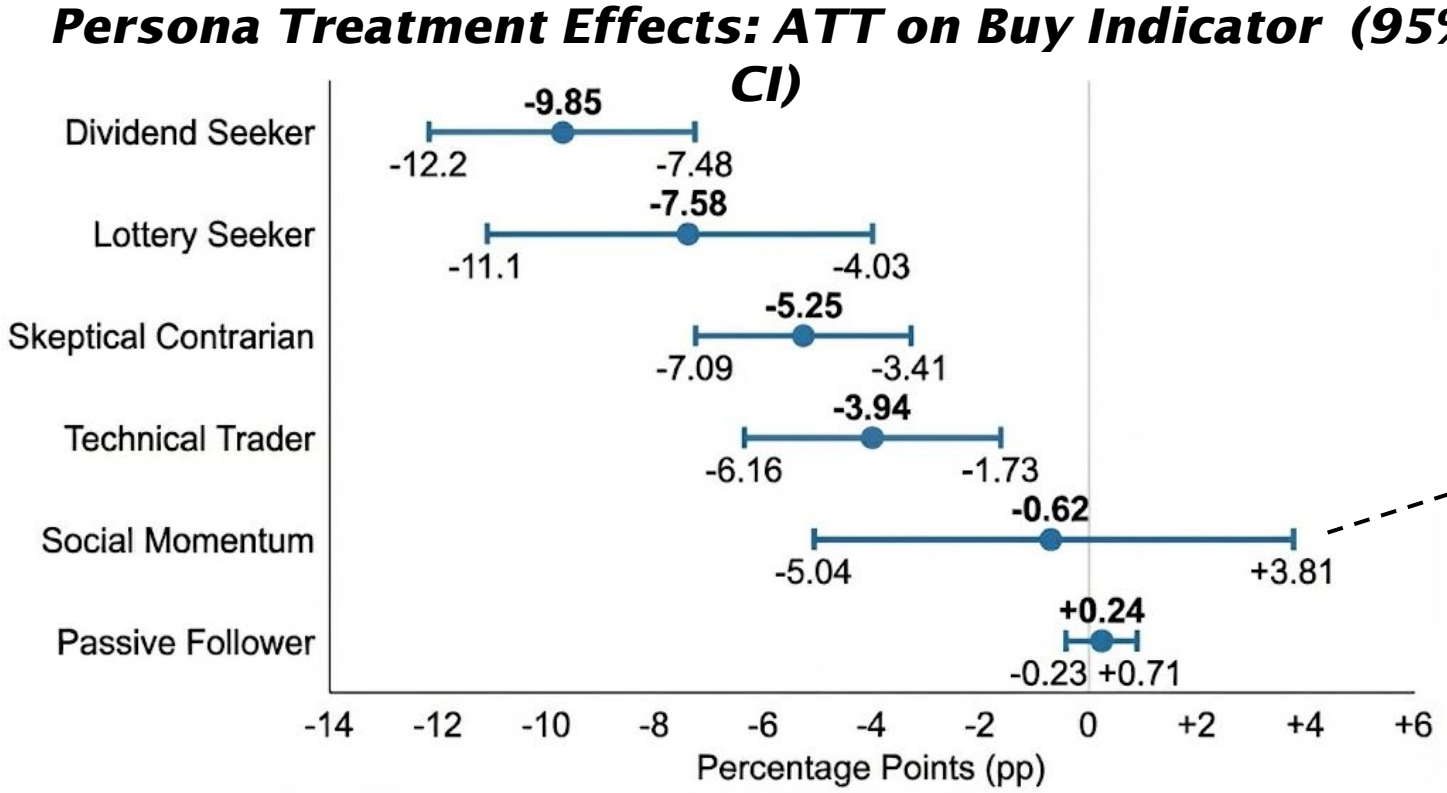
Mechanism & Behavioral Contrast



Attention reduces buying, not selling.

Contradicts human impulse patterns (e.g., FOMO)

The FOMO Paradox:
High-sensitivity personas don't buy the hype



Among Social Momentum (FOMO) agents, 61.1% of HOLD rationales cite a 'negative edge' from the transaction-cost surge (15 bps).

So What?

- Five of six personas show negative effects; Dividend Seeker is most negative (−9.85 pp).
- High-sensitivity Social Momentum shows near-null response, contradicting my hypothesis
- Two channels: skepticism (dominant) + cost sensitivity (moderating)

A Second Reversal: No Loss Aversion

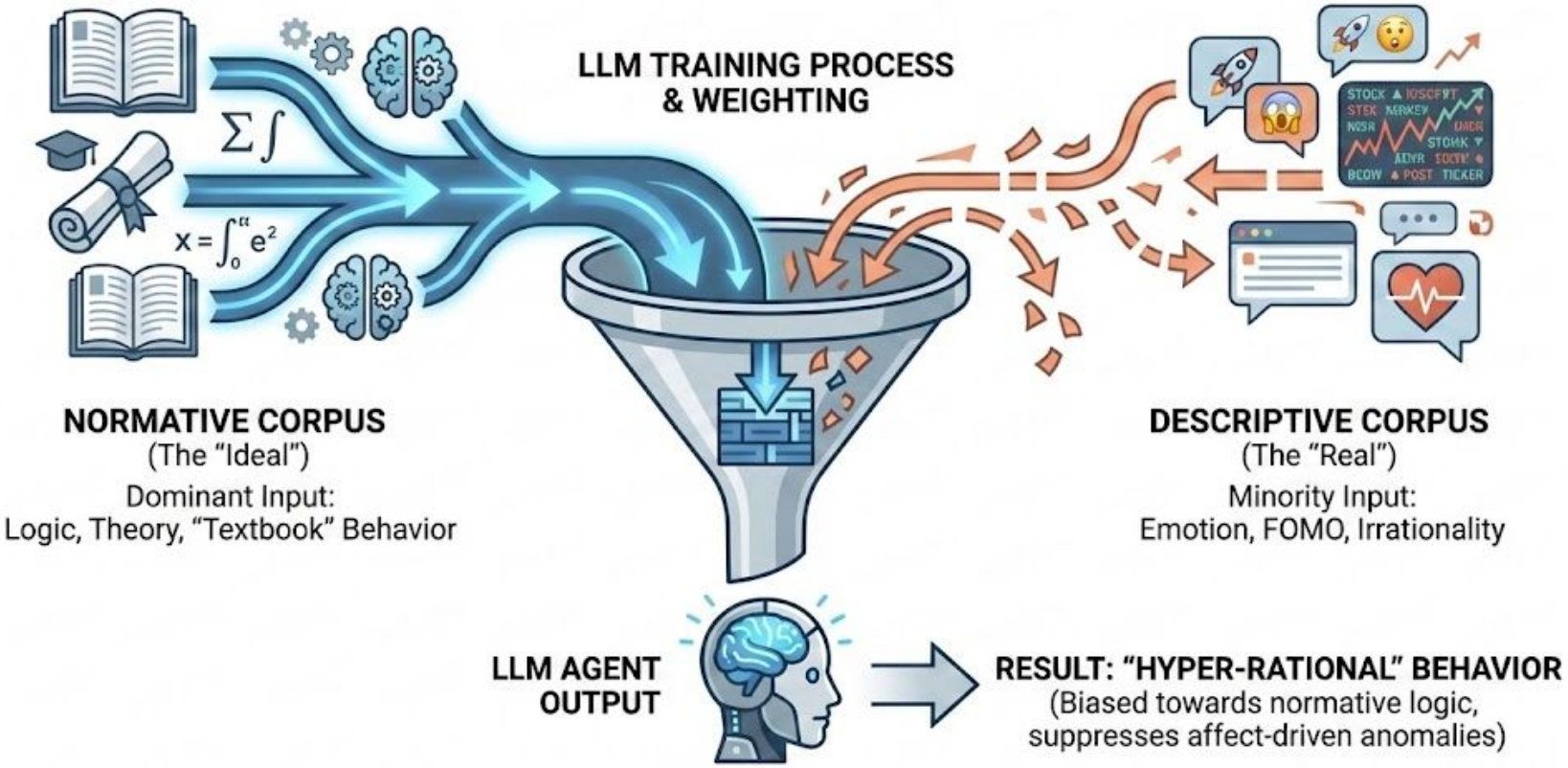
The "disposition effect" is the tendency to sell winning investments too early while holding losing investments too long - a classic human bias driven by loss aversion.

Finding: Agents behave as if running an internal stop-loss strategy (N=3,859 trades).

Result: Attention shock had **no moderating effect**; treated and control groups remain indistinguishable.

Why the Divergence? The NDD Framework

The Normative-Descriptive Divergence (NDD) model explains when LLM agents should diverge from humans. Training data favors normative advice over descriptive behavior.



CAN TEST (Cognitive Mechanisms)

- ☑ Attention allocation
- ☑ Info filtering
- ☑ Rational baselines

CANNOT YET (Affect-Driven Biases)

- ☒ FOMO (Social Momentum)
- ☒ Loss Aversion (Emotional)

LLM Agents vs. Human Investors: A Behavioral Alignment Scorecard

Behavior	Human Pattern	LLM Pattern	Alignment
Attention Allocation	Shifts toward salient assets	Shifts toward salient assets	✅ Aligned
Attention → Trading	Significant buying increase	45% decrease	❌ (Not Aligned)
Disposition Effect	Sell winners, hold losers	Sell losers (3.03x > winners)	🔁 Reversed
Persona Heterogeneity	Sensation-seekers most responsive	Low-sensitivity personas most responsive	🔁 Reversed
Manipulation Compliance	~50%, heterogeneous	99.9%, near perfect	⚠️ Over-Compliant

LLMs simulate cognition (logic) but fail at affection (emotion). Use them to test market mechanisms, not to predict human irrationality. They are simply too smart to be human.

bitly

Paper Link (on SSRN):
bit.ly/497IyU3