# The Limits of AI Trading

Winston W. Dou[◇] Itay Goldstein[◇] Zigang Li[†] Liyan Yang[†]

[◇]University of Pennsylvania and NBER

[†]University of Toronto

NBER Summer Institute – Financial Market Structure

July 2026

# Research question

**When does AI trading outperform humans, and what limits its advantage?**

- **A fundamental but challenging question in AI trading**
  - Requires a clear distinction between AI and human traders
  - Requires humans to anticipate and respond to AI equilibrium behavior

**Our approach:**

- Study a clean, economically meaningful, empirically relevant benchmark:
  - Isolate RL-based trading performance by eliminating AI market power and information advantages (e.g., human-acquired soft information)
- Humans observe private signals and reason strategically, may be bounded
- AI learns from market prices via RL, but has no or weaker private signals

**The paper uses theoretical analysis rather than simulation experiments**

e.g., Dou_Goldstein_Ji (2025, 2026); Colliard_Foucault_Lovo (2026)

- This is more suitable to the research question

# Main findings

1. **The most sophisticated human traders can always outperform AI**
   - AI learns from prices shaped by average, not frontier, human sophistication
   - Humans retain direct access to private information unavailable to AI
2. **Even the least sophisticated human traders can outperform AI**, when private information is either very strong or very weak
   - AI outperforms only when prices are informative enough to guide learning, but human trading is not so aggressive as to eliminate trading rents
3. **Three forces limit AI profitability:**
   - Humans benefit from private information AI traders do not have
   - Sophisticated humans anticipate AI price inference and shade trades
   - **Algorithmic herding** amplifies AI price impact
4. **AI sophistication is shaped by that of humans whose behavior trains it**

# Outline

1. **Trading Environment**

2. **Equilibrium and Algorithmic Herding**

3. **Limits of AI Trading Performance**

# Trading environment

**A single risky asset is traded in a repeated game:**

$$\text{Payoff is } v \sim N(\bar{v}, \sigma_v^2), \text{ and per-capita supply is } S$$

- Normalize $\bar{v} = 0$ and $S = 0$
- Payoffs are i.i.d. over time; omit time subscripts

**Noise-trader demand every period:**

$$z \sim N(0, \sigma_z^2)$$

**Objective-maximizing traders with static objective function:**

$$\pi_i = (v - p)x_i - \frac{\gamma}{2}x_i^2$$

- $p =$ equilibrium asset price
- $x_i =$ investor demand
- $\gamma =$ quadratic trading cost

# Two types of objective-maximizing traders

**Human traders: measure $\psi$**

- Each observes a private signal every period

$$\eta_i = v + e_i, \qquad e_i \sim N(0, \sigma_M^2)$$

- Optimize given their beliefs, but differ in strategic sophistication and thus in how they extract information from prices

**AI traders: measure $\phi$**

- Observe no private signals
- Use only price information and learn through RL
- Algorithms start differently and use different tuning parameters, but all learn from the same market prices

# Two types of human traders

**Cursed investors: measure $\psi_c$**

- Level-0 thinkers who ignore the informational content of prices
- Trade using only their private signal $\eta_i$

**Level-$k$ strategic investors, $k \geq 1$: measure $\psi_s$**

- **Level-$k$ investors who use prices to infer fundamentals**
  ($k \to \infty$ yields fully rational investors)
- Each level-$k$ investor views all others as level-$(k-1)$
- A level-$k$ investor best responds to level-$(k-1)$ strategic behavior
- **Iterated reasoning:** level 0 $\to$ level 1 $\to \cdots \to$ level $k$

**Overall human investors:**

$$\psi_c + \psi_s = \psi$$

# Cursed investors: rational signal use, no price inference

**Cursed (level-0) investors ignore the informational content of prices**

They trade on

$$E[v \mid \eta_i] = \frac{\tau_v \bar{v} + \tau_M \eta_i}{\tau_v + \tau_M}, \qquad \tau_M = \frac{1}{\sigma_M^2}$$

Their demand is

$$x_i^{M,c} = \frac{E[v \mid \eta_i] - p}{\gamma}$$

$$= \underbrace{\frac{\tau_M}{\tau_v + \tau_M}}_{\beta_\eta^{M,0}} \eta_i - \underbrace{\frac{1}{\gamma}}_{\beta_p^{M,0}} p$$

**Interpretation:**

- Cursed investors are not random traders
- They use their own signals correctly
- Their mistake is that they do not learn from prices

# Cursed investors: rational signal use, no price inference

**Cursed (level-0) investors ignore the informational content of prices**

They trade on

$$E[v \mid \eta_i] = \frac{\tau_v \bar{v} + \tau_M \eta_i}{\tau_v + \tau_M}, \qquad \tau_M = \frac{1}{\sigma_M^2}$$

Their demand is

$$\begin{aligned} x_i^{M,c} &= \frac{E[v \mid \eta_i] - p}{\gamma} \\ &= \underbrace{\frac{\tau_M}{\tau_v + \tau_M}}_{\beta_\eta^{M,0}} \eta_i - \underbrace{\frac{1}{\gamma}}_{\beta_p^{M,0}} p \end{aligned}$$

**Interpretation:**

- Cursed investors are not random traders
- They use their own signals correctly
- Their mistake is that they do not learn from prices

# Cursed investors: rational signal use, no price inference

**Cursed (level-0) investors ignore the informational content of prices**

They trade on

$$E[v \mid \eta_i] = \frac{\tau_v \bar{v} + \tau_M \eta_i}{\tau_v + \tau_M}, \qquad \tau_M = \frac{1}{\sigma_M^2}$$

Their demand is

$$\begin{aligned} x_i^{M,c} &= \frac{E[v \mid \eta_i] - p}{\gamma} \\ &= \underbrace{\frac{\tau_M}{\tau_v + \tau_M}}_{\beta_\eta^{M,0}} \eta_i - \underbrace{\frac{1}{\gamma}}_{\beta_p^{M,0}} p \end{aligned}$$

**Interpretation:**

- Cursed investors are not random traders
- They use their own signals correctly
- Their mistake is that they do not learn from prices

Wharton UNIVERSITY of PENNSYLVANIA

# AI traders: no private signals, only prices

**AI traders use stateless Q-learning with Boltzmann exploration**

- No private signals or model of fundamentals and price formation
- Autonomously learn price-contingent demand schedules from profits

**AI trader $j$ chooses a linear demand schedule**

$$x_{j,t}(p_t) = \mu_{j,t} - \beta_{j,t} p_t \in \mathcal{X}$$

and updates its Q-value according to

$$Q_{j,t+1}(x_{j,t}) = (1 - \alpha_{j,t}) Q_{j,t}(x_{j,t}) + \alpha_{j,t} \left[ \pi_{j,t} + \rho \max_{\tilde{x} \in \mathcal{X}} Q_{j,t}(\tilde{x}) \right],$$

where $\alpha_{j,t} \to 0$ governs learning from new trading outcomes

**Boltzmann exploration** selects demand schedule $x \in \mathcal{X}$ with probability

$$q_{j,t}(x) = \frac{\exp\left(Q_{j,t}(x)/\kappa_j\right)}{\sum_{\tilde{x} \in \mathcal{X}} \exp\left(Q_{j,t}(\tilde{x})/\kappa_j\right)},$$

where a higher $\kappa_j$ implies more exploration

# AI traders: no private signals, only prices

**AI traders use stateless Q-learning with Boltzmann exploration**

- No private signals or model of fundamentals and price formation
- Autonomously learn price-contingent demand schedules from profits

**AI trader $j$ chooses a linear demand schedule**

$$x_{j,t}(p_t) = \mu_{j,t} - \beta_{j,t} p_t \in \mathcal{X}$$

and updates its Q-value according to

$$Q_{j,t+1}(x_{j,t}) = (1 - \alpha_{j,t}) Q_{j,t}(x_{j,t}) + \alpha_{j,t} \left[ \pi_{j,t} + \rho \max_{\tilde{x} \in \mathcal{X}} Q_{j,t}(\tilde{x}) \right],$$

where $\alpha_{j,t} \to 0$ governs learning from new trading outcomes

**Boltzmann exploration** selects demand schedule $x \in \mathcal{X}$ with probability

$$q_{j,t}(x) = \frac{\exp\left(Q_{j,t}(x)/\kappa_j\right)}{\sum_{\tilde{x} \in \mathcal{X}} \exp\left(Q_{j,t}(\tilde{x})/\kappa_j\right)},$$

where a higher $\kappa_j$ implies more exploration

# AI traders: no private signals, only prices

**AI traders use stateless Q-learning with Boltzmann exploration**

- No private signals or model of fundamentals and price formation
- Autonomously learn price-contingent demand schedules from profits

**AI trader $j$ chooses a linear demand schedule**

$$x_{j,t}(p_t) = \mu_{j,t} - \beta_{j,t} p_t \in \mathcal{X}$$

and updates its Q-value according to

$$Q_{j,t+1}(x_{j,t}) = (1 - \alpha_{j,t}) Q_{j,t}(x_{j,t}) + \alpha_{j,t} \left[ \pi_{j,t} + \rho \max_{\tilde{x} \in \mathcal{X}} Q_{j,t}(\tilde{x}) \right],$$

where $\alpha_{j,t} \to 0$ governs learning from new trading outcomes

**Boltzmann exploration** selects demand schedule $x \in \mathcal{X}$ with probability

$$q_{j,t}(x) = \frac{\exp\big(Q_{j,t}(x)/\kappa_j\big)}{\sum_{\tilde{x} \in \mathcal{X}} \exp\big(Q_{j,t}(\tilde{x})/\kappa_j\big)},$$

where a higher $\kappa_j$ implies more exploration

# AI learning, convergence, & herding

**Let $Q^{(k)}$ denote the limiting Q-function learned in the environment with level-$k$ investors**

**Algorithmic herding:** AI traders converge to the optimal strategy of uninformed rational-expectations investors in the environment with level-$k$ investors

$$x^{A,k} = \frac{E_k[v \mid p] - p}{\gamma} = -\beta_p^{A,k} p + \bar{x}^{A,k}$$

where $E_k[\cdot \mid p]$ captures the rational expectation based on equilibrium price $p$ in the environment with level-$k$ investors

**AI traders are smart, but their training is only as good as the information content in the price**

# AI learning, convergence, & herding

**Let $Q^{(k)}$ denote the limiting Q-function learned in the environment with level-$k$ investors**

**Algorithmic herding:** AI traders converge to the optimal strategy of uninformed rational-expectations investors in the environment with level-$k$ investors

$$x^{A,k} = \frac{E_k[v \mid p] - p}{\gamma} = -\beta_p^{A,k} p + \bar{x}^{A,k}$$

where $E_k[\cdot \mid p]$ captures the rational expectation based on equilibrium price $p$ in the environment with level-$k$ investors

**AI traders are smart, but their training is only as good as the information content in the price**

Wharton
University of Pennsylvania

# AI learning, convergence, & herding

**Let $Q^{(k)}$ denote the limiting Q-function learned in the environment with level-$k$ investors**

**Algorithmic herding:** AI traders converge to the optimal strategy of uninformed rational-expectations investors in the environment with level-$k$ investors

$$x^{A,k} = \frac{E_k[v \mid p] - p}{\gamma} = -\beta_p^{A,k} p + \bar{x}^{A,k}$$

where $E_k[\cdot \mid p]$ captures the rational expectation based on equilibrium price $p$ in the environment with level-$k$ investors

**AI traders are smart, but their training is only as good as the information content in the price**

# The price as an endogenous signal for AI

In the objective environment, **the equilibrium price is linear**:

$$p = \bar{p} + \frac{\zeta}{\xi} v + \frac{1}{\xi} z$$

- $\bar{p} =$ a constant price shifter that does not affect price informativeness
- $\zeta =$ fundamental loading in price
- $\xi =$ market depth coefficient
- $z =$ noise-trader demand

**Price informativeness** is summarized by

$$\boldsymbol{\mathcal{I}_p \equiv \zeta^2 \tau_z}, \quad \text{with} \quad \tau_z = \frac{1}{\sigma_z^2}$$

**High $\mathcal{I}_p \Leftrightarrow$ price is a precise signal of fundamentals**

# Level-$k$ strategic investors

**Level-0 investors are cursed investors**

$$x_i^{M,c} = x_i^{M,0} = \frac{E[v \mid \eta_i] - p}{\gamma} = \beta_\eta^{M,0}\eta_i - \beta_p^{M,0}p$$

**For each $k = 1, 2, \ldots$, the level-$k$ reasoning proceeds in two steps:**

- **Perception:** Each level-$k$ investor believes the market consists of
  - Cursed with measure $\psi_c$
  - Level-$(k-1)$ investors with measure $\psi_s$
  - AI traders with measure $\phi$, whose converged strategy is $Q^{(k-1)}$
- **Best response:** Each level-$k$ investor chooses optimal demand using
  - Her private signal $\eta_i$
  - The rational expectation $E_{k-1}[\cdot \mid p]$ based on the perceived equilibrium price $p$ in the environment with level-$(k-1)$ investors

$$x_i^{M,k} = \frac{E_{k-1}[v \mid \eta_i, p] - p}{\gamma} = \beta_\eta^{M,k}\eta_i - \beta_p^{M,k}p + \bar{x}^{M,k}$$

# Level-$k$ strategic investors

**Level-0 investors are cursed investors**

$$x_i^{M,c} = x_i^{M,0} = \frac{E[v \mid \eta_i] - p}{\gamma} = \beta_\eta^{M,0}\eta_i - \beta_p^{M,0}p$$

**For each $k = 1, 2, \ldots$, the level-$k$ reasoning proceeds in two steps:**

- **Perception:** Each level-$k$ investor believes the market consists of
  - Cursed with measure $\psi_c$
  - Level-$(k-1)$ investors with measure $\psi_s$
  - AI traders with measure $\phi$, whose converged strategy is $Q^{(k-1)}$
- **Best response:** Each level-$k$ investor chooses optimal demand using
  - Her private signal $\eta_i$
  - The rational expectation $E_{k-1}[\cdot \mid p]$ based on the perceived equilibrium price $p$ in the environment with level-$(k-1)$ investors

$$x_i^{M,k} = \frac{E_{k-1}[v \mid \eta_i, p] - p}{\gamma} = \beta_\eta^{M,k}\eta_i - \beta_p^{M,k}p + \bar{x}^{M,k}$$

# Level-$k$ strategic investors

**Level-0 investors are cursed investors**

$$x_i^{M,c} = x_i^{M,0} = \frac{E[v \mid \eta_i] - p}{\gamma} = \beta_\eta^{M,0}\eta_i - \beta_p^{M,0} p$$

**For each $k = 1, 2, \ldots$, the level-$k$ reasoning proceeds in two steps:**

- **Perception:** Each level-$k$ investor believes the market consists of
  - Cursed with measure $\psi_c$
  - Level-$(k-1)$ investors with measure $\psi_s$
  - AI traders with measure $\phi$, whose converged strategy is $Q^{(k-1)}$
- **Best response:** Each level-$k$ investor chooses optimal demand using
  - Her private signal $\eta_i$
  - The rational expectation $E_{k-1}[\cdot \mid p]$ based on the perceived equilibrium price $p$ in the environment with level-$(k-1)$ investors

$$x_i^{M,k} = \frac{E_{k-1}[v \mid \eta_i, p] - p}{\gamma} = \beta_\eta^{M,k}\eta_i - \beta_p^{M,k} p + \bar{x}^{M,k}$$

# Level-$k$ strategic investors

**Level-0 investors are cursed investors**

$$x_i^{M,c} = x_i^{M,0} = \frac{E[v \mid \eta_i] - p}{\gamma} = \beta_\eta^{M,0}\eta_i - \beta_p^{M,0}p$$

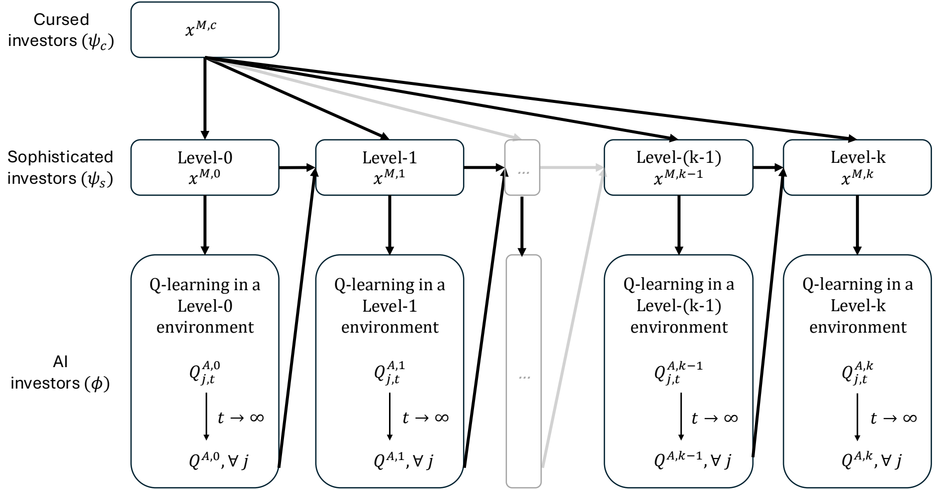
**For each $k = 1, 2, \ldots$, the level-$k$ reasoning proceeds in two steps:**

- **Perception:** Each level-$k$ investor believes the market consists of
  - Cursed with measure $\psi_c$
  - Level-$(k-1)$ investors with measure $\psi_s$
  - AI traders with measure $\phi$, whose converged strategy is $Q^{(k-1)}$
- **Best response:** Each level-$k$ investor chooses optimal demand using
  - Her private signal $\eta_i$
  - The rational expectation $E_{k-1}[\cdot \mid p]$ based on the perceived equilibrium price $p$ in the environment with level-$(k-1)$ investors

$$x_i^{M,k} = \frac{E_{k-1}[v \mid \eta_i, p] - p}{\gamma} = \beta_\eta^{M,k}\eta_i - \beta_p^{M,k}p + \bar{x}^{M,k}$$

# Illustration: Level-$k$ iteration with AI traders

## Bounded-rational anticipation of learned AI trading



# Outline

1. **Trading Environment**

2. **Equilibrium and Algorithmic Herding**

3. **Limits of AI Trading Performance**

# Equilibrium: Market clearing and investors' profits

**Market price $p$ is the market-clearing price**

- The market price clears total demand from cursed investors, sophisticated investors, AI traders, and noise traders

**Cursed investors**

- Earn profits from private signals

**Level-$k$ investors**

- Earn profits from combining private signals with price information
- Higher profits under higher sophistication

**AI traders**

- Learn from equilibrium prices, not private signals
- Perform well when prices are informative enough to guide learning

# Algorithmic herding in equilibrium

**Theorem 1: AI traders using Q-learning with Boltzmann exploration converge to the behavior of uninformed rational-expectations investors**

At iteration 0: AI traders start with heterogeneous policies and fixed algorithmic parameters

At iteration $t$, each AI trader $i$ chooses a linear demand rule:

$$x_{i,t}^{A,k} = -\beta_{i,t}^{A,k} p + \bar{x}_{i,t}^{A,k}, \quad \text{for every given } k$$

Under general regularity conditions: as $t \to \infty$, for all $i$

$$E[\beta_{i,t}^{A,k}] \to \beta^{A,k} \quad \text{and} \quad E[\bar{x}_{i,t}^{A,k}] \to \bar{x}^{A,k}, \quad \text{for every given } k,$$

where $(\beta^{A,k}, \bar{x}^{A,k})$ is the optimal demand of uninformed rational-expectations investors in the environment with level-$k$ investors

# What ensures convergence (mean field equilibrium)?

**Convergence comes from stable learning plus weak equilibrium feedback**

- **Individual learning is stable:** the Q-learning step size declines slowly enough to keep learning, but fast enough for noise to vanish
- **Exploration is sufficient:** Boltzmann exploration (temperature $\tau$) gives every trading rule positive probability, so expected rewards can be learned
- **Aggregate feedback is weak:** changes in population behavior have only limited effects on prices and rewards
    - Formally, the reward-policy feedback is a contraction:

$$\frac{4\theta}{\tau} < 1$$

    where $\theta$ measures how sensitive a trading rule's expected reward is to aggregate AI behavior

**Together, these conditions ensure that learning converges as in a stable single-agent system**, even though it is embedded in an equilibrium market

# Strategy crowding in equilibrium

**Theorem 2: As more AI traders deploy the same learned price-contingent strategy ($\phi \to \infty$), per-investor AI profits decline and converge to zero**

- Algorithmic herding creates a sector-level crowding externality
- Each AI trader is atomistic and takes prices as given, but a positive mass of investors deploying the same learned rule changes the equilibrium price process faced by all AI traders, compressing their trading rents

# Outline

1. **Trading Environment**
2. **Equilibrium and Algorithmic Herding**
3. **Limits of AI Trading Performance**

# Theorem: Level-$\infty$ investors always outperform AI

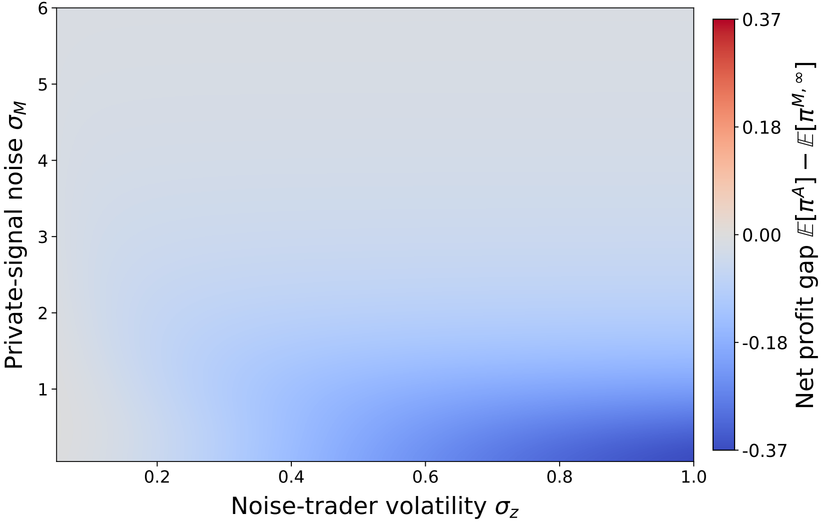
In an environment with level-$\infty$ investors, the sophisticated (level-$\infty$) trader earns strictly higher expected trading profits than the AI trader:

$$\Pi^{M,\infty} > \Pi^{A}$$

**Intuition:**

- Level-$\infty$ investors combine private signals with price information
- AI learns only from prices that are noisy summaries of investor information
- Sophisticated investors shade their trades, reducing what AI can exploit

# Illustration: $\Pi^A - \Pi^{M,\infty}$




Note: $\psi_c = 0.3$, $\psi_s = 0.3$, $\phi = 0.4$, $\gamma = 1$, and $\sigma_v = 1$

# Theorem: Cursed investors win with large noise

**If noise trading $\sigma_z$ is sufficiently large:**

$$\sigma_z > \frac{\psi \sigma_v}{2\gamma},$$

then, in an environment with level-$\infty$ investors, the cursed investor earns strictly higher expected trading profits than the AI trader:

$$\Pi^{M,c} > \Pi^{A}$$

**Intuition**

- When noise trading is large, prices are too noisy for AI traders
- Price precision remains below private-signal precision
- AI learns from prices, but prices are always a noisier signal than rational investors' private information

# Theorem: Small noise makes AI dominate sometimes

**When noise trading $\sigma_z$ is sufficiently small:**

$$\sigma_z < \frac{\psi_c \sigma_v}{2\gamma},$$

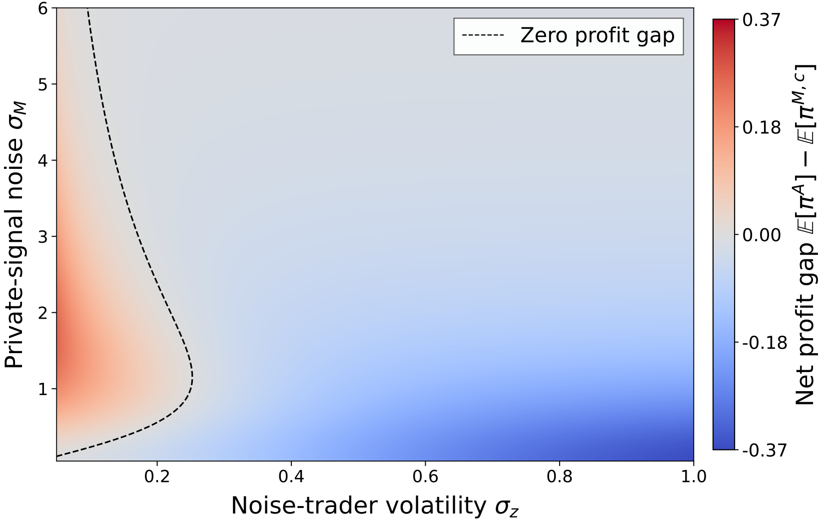
then, there exist thresholds

$$\sigma_M^L(\sigma_z) < \sigma_M^H(\sigma_z)$$

such that the cursed investor earns lower expected trading profits than the AI trader if and only if $\sigma_M \in \left[\sigma_M^L(\sigma_z), \sigma_M^H(\sigma_z)\right]$:

$$\Pi^{M,c} < \Pi^A \quad \Longleftrightarrow \quad \sigma_M \in \left[\sigma_M^L(\sigma_z), \sigma_M^H(\sigma_z)\right]$$

# Illustration: $\Pi^A - \Pi^{M,c}$




Note: $\psi_c = 0.3$, $\psi_s = 0.3$, $\phi = 0.4$, $\gamma = 1$, and $\sigma_v = 1$

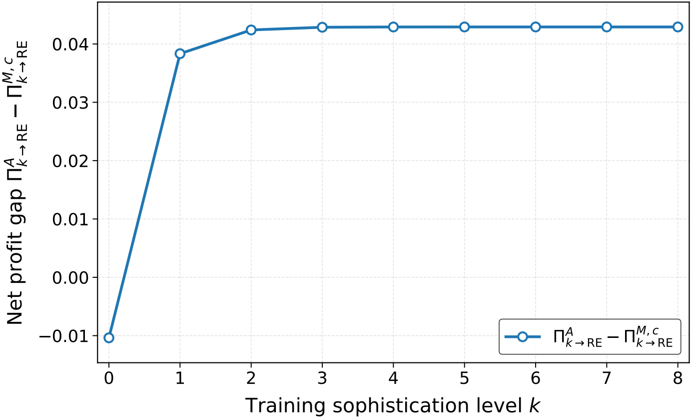
# Intuition: A narrow AI-dominance window

**AI performance relative to cursed investors is non-monotonic in private-signal precision**

- Very precise private signals: cursed investors' direct information dominates AI's noisy price signal
- Very imprecise private signals: too little fundamental information enters prices for AI to learn
- Intermediate precision and low noise trading: prices aggregate enough information for AI to outperform cursed investors

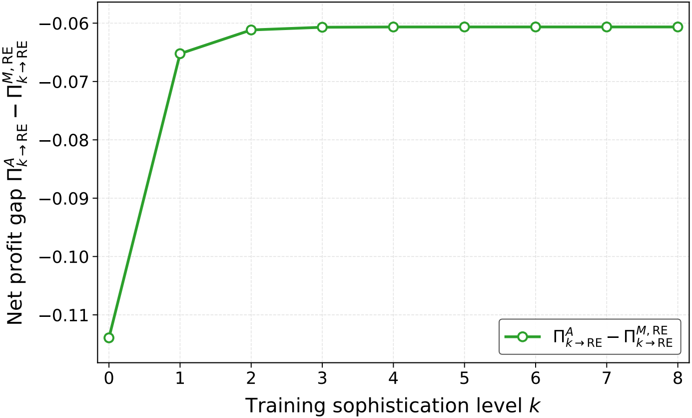
**AI dominance is confined to a narrow window**

# Theorem: Human sophistication $\Rightarrow$ AI sophistication




Note: $\sigma_M = 1$, $\sigma_z = 0.2$, $\psi_c = 0.3$, $\psi_s = 0.3$, $\phi = 0.4$, $\gamma = 1$, and $\sigma_v = 1$

# Theorem: Human sophistication $\Rightarrow$ AI sophistication




Note: $\sigma_M = 1$, $\sigma_z = 0.2$, $\psi_c = 0.3$, $\psi_s = 0.3$, $\phi = 0.4$, $\gamma = 1$, and $\sigma_v = 1$

# Conclusions

**1. The most sophisticated human traders can always outperform AI**

- AI learns from prices shaped by average, not frontier, human sophistication
- Humans retain direct access to private information unavailable to AI

**2. Even the least sophisticated human traders can outperform AI**, when private information is either very strong or very weak

- AI outperforms only when prices are informative enough to guide learning, but human trading is not so aggressive as to eliminate trading rents

**3. Three forces limit AI profitability:**

- Humans benefit from private information AI traders do not have
- Sophisticated humans anticipate AI price inference and shade trades
- **Algorithmic herding** amplifies AI price impact

**4. AI sophistication is shaped by that of humans whose behavior trains it**