AI TRADING IN FINANCIAL MARKETS

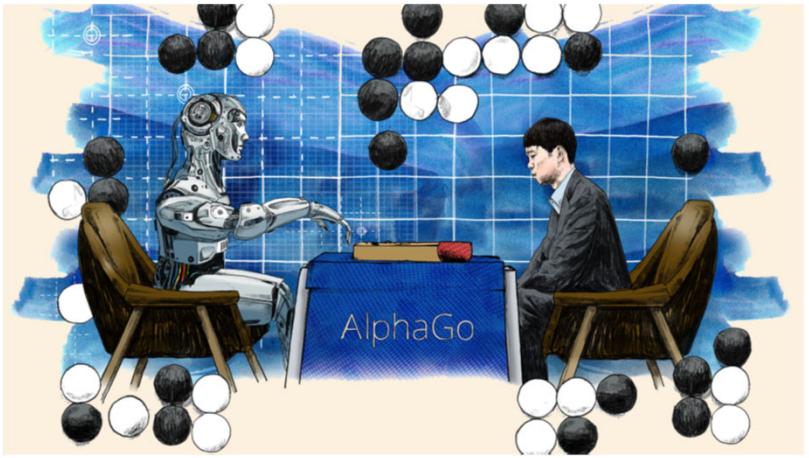
PANEL ON AI AND FINANCIAL STABILITY

ECB FORUM ON CENTRAL BANKING 2026

ITAY GOLDSTEIN, WHARTON

WHAT IS "AI-POWERED TRADING?"

- AI-powered trading:
  - Algorithmic trading system + reinforcement-learning algorithms
  - Reinforcement learning is a key approach of AI, and serves as the backbone of many famous applications such as "AlphaGo"
- Implementation: Q-Learning
  - AI agent keeps a Q-value table, summarizing expected payoff for every pair of state and action
  - After each step, the Q-value of a particular pair is updated by blending new experience with previous knowledge
  - The AI agent chooses actions alternating between exploitation and exploration
  - Over time, the AI agent discovers effective strategies on its own, without needing prior knowledge of the environment



RISKS TO FINANCIAL MARKETS

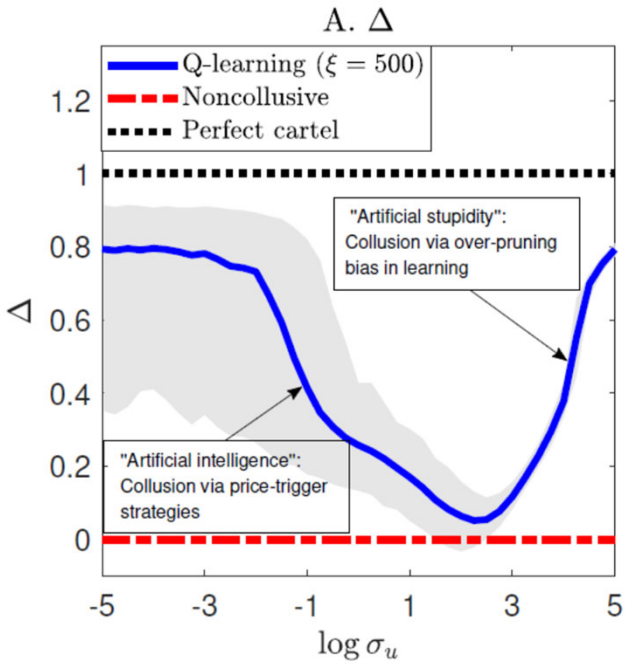
- When evaluating risks for financial markets, key questions are:
  - How do AI trading algorithms interact with each other?
  - What does this imply for financial market outcomes?
- The approach taken with human traders may not be sufficient
  - Growing recognition that artificial intelligence and human intelligence are different
  - Legal/regulatory treatment for human players might not be suitable for AI
    - Detection of human collusion is based on explicit communication or shared intent, but AI may reach these outcomes differently

EXPERIMENTAL RESEARCH APPROACH

- Traditional research approaches pose challenges
  - Multi-agent systems with reinforcement learning do not lend themselves to theoretical analysis
  - Empirical analysis is not readily available given data limitations
- A path forward: Experimental studies
  - But with AI algorithms, not humans, as the economic agents
- The idea is to simulate the behavior of AI trading algorithms in a financial market environment and see what patterns they develop
- Relative to human experiments, AI experiments are:
  - Easier to do at large scale
  - Less subject to external validity concerns

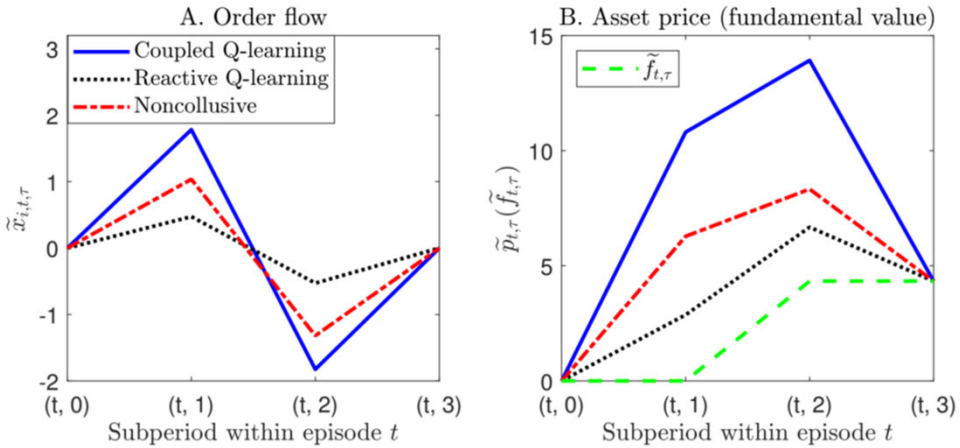
COLLUSION:

- Dou, Goldstein, Ji (2025a):
  - AI algorithms acting as speculators in financial markets learn to trade less aggressively against each other making supra-competitive profits
- Figure on the right shows profits as a function of noise in the market
  - With small noise, collusion is achieved via price-trigger strategies (“artificial intelligence”)
  - With large noise, collusion is achieved via over-pruning learning bias (“artificial stupidity”)
- When collusion arises, market quality is harmed:
  - Lower liquidity, lower price informativeness



MANIPULATION:

- Dou, Goldstein, Ji (2025b):
  - AI algorithms acting as speculators go beyond "reactive Q-learning" into "AI planning" or "coupled Q-learning"
  - They learn to coordinate on setting price paths that benefit them
- Figure on the right shows how they:
  - Create bubbles by trading aggressively in tandem
  - Lead to crash-like reversal by coordinated unwinding of their positions
- Under this kind of behavior:
  - Markets are exposed to greater fragility and prices exhibit higher volatility



CONCLUSIONS

- We need to understand how markets might behave following the introduction of AI trading algorithms
  - This is about the "psychology" of AI traders and how they interact
- Experimental approach is readily available for analyzing such issues
  - It doesn't suffer as much from the external validity problem of human experiments
- Analysis shows patterns of collusion and manipulation
  - This raises questions about market regulation in the world of AI traders, where explicit communication is not the culprit
- Finally, we need to understand how humans and AI algorithms trade against each other
  - Initial theoretical analysis suggests that humans may still have an advantage in many cases