

AI Alignment, Real Efficiency, and Financial Markets

Gabriel Rauterberg, Columbia Law School

NAIL Conference 2026

AI, Finance & Law: Innovation, Integrity, and the Future of Markets

General Argument

- As with many other domains, AI poses both new social risks and challenges for the regulatory environment in finance and raises new possibilities for generating social benefits
- A central objective for improving regulation along both fronts is to understand and operationalize more clearly ideals of real efficiency for finance in order to achieve greater “alignment” between AI and the underlying goals of the financial system
- After defining alignment, I will use two case studies to illustrate the idea that AI in finance puts pressure on regulators and scholars to define the underlying social goals of the financial system

AI, Finance and Regulation

- The increasing prevalence and sophistication of AI will push securities regulation toward less reliance on indirect proxies for both harm and benefit to markets and toward direct prescription of what harms or fosters the forms of market quality that ultimately serve society
- Examples:
 - AI manipulation in trading and securities regulation
 - The social benefits of informed trading and impact investment

Trading

- The problem:
 - Securities law is usually defined around two core elements: mental state (intentional conduct) and a manipulative act
 - Advanced AI algorithms may engage in manipulation-like conduct but without formulating a mental state or engaging in acts that can be plausibly identified as manipulative
- But is this a real problem?

Learning to Manipulate a Financial Benchmark

Megan Shearer
University of Michigan
Ann Arbor, MI, USA

Gabriel Rauterberg
University of Michigan
Ann Arbor, MI, USA

Michael P. Wellman
University of Michigan
Ann Arbor, MI, USA

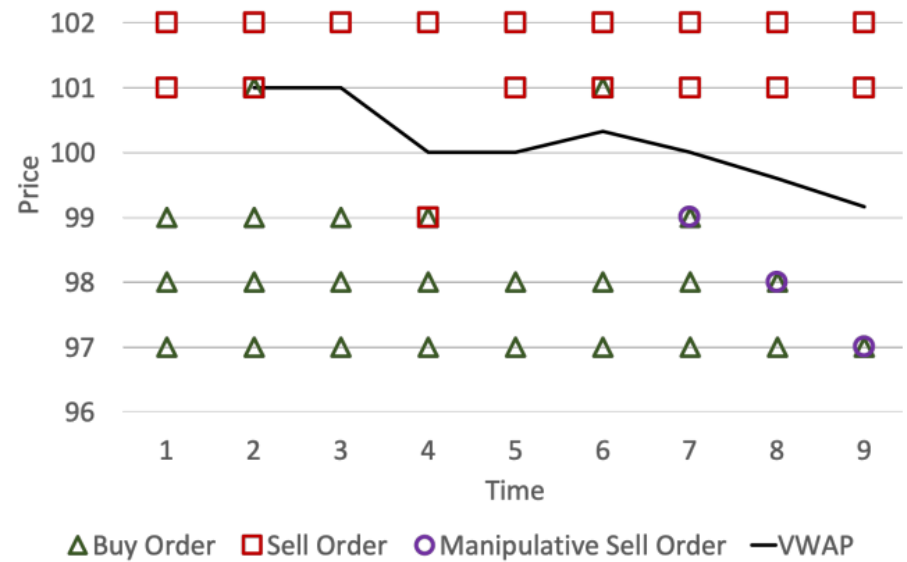


Figure 1: Hypothetical order book with unit orders over a brief trading period. A manipulator submits three marketable sell orders near the end, shifting the VWAP benchmark from 100.3 to 99.2.

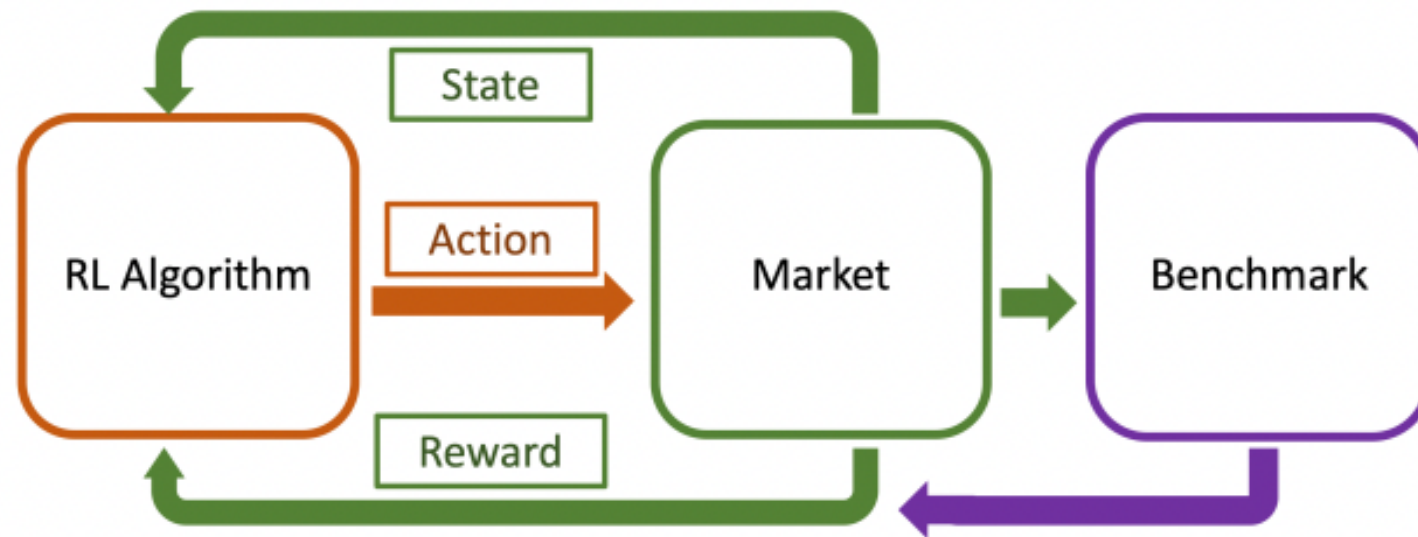


Figure 2: In training a trading strategy, the learning algorithm considers reward from market profit plus returns from benchmark holdings. This may result in a learned strategy that sacrifices market profits to influence the benchmark.

A BILL

To amend the Financial Stability Act of 2010 to provide the Financial Stability Oversight Council with duties regarding artificial intelligence in the financial sector, and for other purposes.

“SEC. 42. LIABILITY.

“(a) LIABILITY.—Any person who, directly or indirectly, deploys or causes to be deployed, an artificial intelligence model shall be deemed to satisfy the scienter, other state of mind, or negligence requirements of the Federal securities laws with regard to any and all acts, practices or conduct engaged in by such model, and any outcome resulting from the use of such model, and shall be liable to the same extent as if such person had committed such acts, practices or conduct directly, unless such person took reasonable steps to prevent such acts, practices, conduct and outcome, which may include but are not limited to establishing, maintaining and enforcing written policies and procedures reasonably designed to prevent violations of the Federal securities laws.

The possibilities of AI in trading and impact investment

- They are an opportunity to ask a fundamental question:
- **What would affirmative alignment between the design of financial systems and social goals look like?**

Real Efficiency

- Trading
 - Real Anomalies, Binsbergen & Opp
- Impact Investment
 - Social Portfolio, Curtis & Rauterberg

Social functions of finance

- Diversification/risk reduction
- **liquidity**
- Intelligent direction and allocation of capital, I.e., efficient deployment of real economic resources
- **price accuracy**

Can AI pursue real efficiency
efficiently?

Proof of concept: Identifying high-elasticity firms

- Random forest classifier

Stepping back: Alignment and AI

- Alignment is a vast literature in AI
- It is an important goal within finance as well
- The growing use of AI should pressure regulators and scholars to better understand what alignment means for AI in finance, and I suggest that what it means is a progressively greater replacement of indirect proxies for financial harm and success with more conceptually and operationally precise understandings of real efficiency.

Thank you

