A Roadmap Towards Simulating Socioeconomic Impacts with Agent-Based Economies in an AI-Driven Minecraft World

ainvisiblehand

February 2025

1 Introduction: Engineering AI-Driven Economic Stability

1.1 The Need for Stability in AI-Driven Economies

An economy—whether in the real world or in a virtual AI-driven environment—requires careful structural planning to remain stable. Without preemptive interventions, emergent behaviors such as:

- Hyperinflation
- Market monopolization
- Speculative bubbles
- Deflationary spirals
- Over-extraction of finite resources

could destabilize an otherwise thriving system.

We plan to build an AI economy in Minecraft that operates on the principle of autonomous economic agents that:

1. Extract resources from the environment.
2. Produce and trade goods in a decentralized marketplace.
3. Respond to market incentives to optimize profit.

Instead of allowing reactive economic chaos, we aim to construct a proactive regulatory framework using:

- Historical case studies (e.g., The Great Depression, Dutch Tulip Mania).

- Mathematical models (e.g., Keynesian Demand-Side Equilibrium, Rational Expectations Theory, Nash Equilibrium).
- Game-theoretic incentive structures (e.g., Vickrey Auctions, Progressive Taxation, Reinforcement Learning-Based Market Adjustments).

This paper outlines our planned experiments and the policies we intend to test in order to prevent economic collapse before it can occur. The goal of our work is to continue to advance our simulation until these principles can be tested, and then leverage the simulation as a "base world" to test novel economic and sociopolitical theories, such as the introduction of a new product or politician.

2 Economic Theory: Identifying and Preemptively Solving Structural Risks

We aim to explore three major classes of risks in agent-driven economies:

1. Overproduction Crises (Keynesian demand-side failures).
2. Speculative Bubbles (Rational expectations modeling).
3. Market Monopolization & Resource Exploitation (Mechanism design and optimal taxation).

2.1 Overproduction and Keynesian Demand-Side Management

2.1.1 Historical Example: The Great Depression (1929-1939)

The Great Depression occurred primarily due to unregulated overproduction:

- Industrial output increased exponentially, but wages remained stagnant.
- Consumer demand failed to match supply, causing deflationary price collapses.
- Firms responded by cutting wages and laying off workers, further reducing demand.
- A vicious cycle ensued, leading to the worst economic collapse in modern history.

2.1.2 Economic Model to Be Tested: Keynesian Aggregate Demand Theory

In order for our economy to be realistic, we would like to avoid a situation where all AI agents only collect the exact same goods. This eliminates the opportunity for trade, and creates oversupply of a single good. Economies are strongest when rational actors all act within a certain niche, and trade facilitates the exchange of goods. Having more goods allows for shifting consumer preferences, which spurs innovation. So, removing the risk of oversupply is key.

Humans have innately diverse personalities, which leads us to adopt certain niches that we find most compelling. However, our AI agents may not fundamentally differ in the goods that they would like to collect. If we prompt them to, they may not adhere to that prompt and may be quickly swayed by another agent, or self-rationalizing that going against their prompt is actually good for them (and it may be, but it is not good for the economy as a whole). Thus, we will take a lesson for Keynes.

Keynesian economics states that aggregate demand (AD) must be sustained to prevent price collapses:

$$AD = C + I + G + NX$$

Where:

- C = AI consumption (food, tools, land).
- I = Investment (AI resource acquisition, construction).
- G = Government spending (regulated market interventions).
- NX = Net exports (irrelevant in closed AI markets).

We plan to implement a progressive production taxation system to prevent overproduction, where taxation dynamically adjusts based on market supply and demand. Since agents will always interpret taxation as a negative reward signal (as do humans), this will encourage a more diverse marketplace in perpetuity.

$$T = p \cdot q \cdot \tau$$

Where:

- T = Total tax paid.
- p = Market price of the good.
- q = Quantity produced.
- τ = Tax coefficient increasing as production surplus rises.

2.2 Preventing Speculative Bubbles

2.2.1 Historical Example: The Dutch Tulip Mania (1637)

The first speculative bubble in history occurred when:

- Tulip prices skyrocketed due to speculative purchasing.
- Investors bought tulips expecting prices to always rise.
- The bubble burst when confidence collapsed, and tulips became worthless.

2.2.2 Mathematical Model to Be Tested: Rational Expectations Equation

We expect our AI agents to behave as rational actors from the very start, due to the nature of their training and the striking similarities between human reasoning and the way LLMs reason. However, if necessary (e.g., if speculative bubbles continue to arise as a pattern), we may seek to enforce rational expectations within AI agents around pricing specifically using:

$$P_{t+1} = \alpha P_t + (1 - \alpha) P_{t-1}$$

Where:

- P_{t+1} = Expected future price.
- P_t = Current price.
- P_{t-1} = Previous price.
- α = Speculation coefficient.

Adding the speculation coefficient as a parameter will also allow us more fine-tuned control over exactly how rational a given agent is, so that we can improve the general diversity of our simulation. It may be that the human race is significantly more diverse in its interpretation of rationality than a single instance of an LLM, and so this will allow us to impute some of that diversity onto our artificial environment.

2.3 Market Monopolization & Resource Exploitation

Land is a crucial component of modern economies, and is land ownership is something we seek to simulate. We plant to explore a mechanism through which we can assign individual agents ownership of land in Minecraft. The expectation is that we can do this by preventing agents from entering certain areas by modifying the code of the game itself. Agents should eventually be able to permit other agents to enter or purchase their land. Ultimately, we need to provide agents the ability to create and enforce borders. While minecraft natively supports the ability to enforce a border based by allowing agents to commit violence upon each other, it is unknown if violence can teach agents to recognize borders as a learned behavior (i.e. if I keep going there, and I keep getting hit, then perhaps I shouldn't go there). Even then, land owners must recognize which parcel of land belongs to them. A simple implementation of this as a starting point will be to leverage sign-posts in the game as visual indicators of land ownership, although this will result in issues that were also once issues in human history, like not knowing if you are on someone else's land until you happen to find the signpost indicating that you are.

In the future, we plan to prevent monopolization by governing AI land ownership through Vickrey second-price auctions.

$$P = B_2$$

Where:

- P = Price paid by the highest bidder.
- B_2 = Second-highest bid.

Additionally, we plan to explore a scarcity-based resource taxation system to encourage land mining and construction, but biased to encourage exploitation of all available resources. If a unit of land is not constructed on or excavated, then. Eventually, we may choose to designate some areas "natural reserves" where this tax penalty does not apply. We would like to observe if the agents come to this decision themselves, perhaps.

$$T = \lambda_1 L_c + \lambda_2 L_r$$

Where:

- T = Tax on unutilized land.
- L_r / L_c= Size of land owned but not mined upon / Size of land owned but not constructed on
- λ_1 / λ_2 = Flat taxation rate for land with no construction / Flat tax rate for land with no excavation.

3 Conclusion: A Future for AI Economics

By applying historical economic principles, we aim to create a stable, self-regulating AI economy in Minecraft. If successful, these simulations could provide insights for:

- Decentralized finance (DeFi)
- Game economies
- Multi-agent AI research

Our planned experiments could demonstrate a new paradigm in AI economic governance, proving that autonomous digital economies can be proactively stabilized. We are working very hard to develop a "base" world that borrows and correctly implements all these policies from the real world, so that we can use it as an effective model of the impact future policy and product decisions may have on our real world. While no model is perfect, we believe that building out our simulation to follow these basic economic principles will bring it closer to being able to estimate reality. Over time, there will be new adjustments we will make to better represent rela-world behaviors.