

A Recursively-Inspired Framework and Simulation Results for Evaluating a Value-Based UBI Policy

Ram Potham

April 27, 2025

Abstract

The potential for Transformative AI (TAI) to exacerbate inequality necessitates exploring novel redistributive policies. This document presents a comprehensive framework and reports results from a conceptual simulation study evaluating the economic implications of a Universal Basic Income (UBI) funded by a significant annual tax (e.g., 4%) on company value. Drawing insights from recursive alignment (RA) and functional models of intelligence (FMI), this framework emphasizes policy adaptability and potential failure modes arising from rigid design under TAI's deep uncertainty. This document details the core research questions, analytical focal points, conceptual model structure, RA-inspired simulation design, and interprets the simulation outputs. It serves as a proof-of-concept and guide for future research into designing resilient economic policies for the TAI era.

Contents

1	Introduction	3
2	Research Question and Core Focus	3
3	Methodology: RA-Inspired Framework and Simulation	4
3.1	Conceptual Framework	4
3.2	Simulation Design Blueprint	4
4	Simulation Execution and Results	5
4.1	Simulation Execution	5
4.2	Simulation Results	5
5	Interpretation and Discussion of Simulation Results	5
5.1	Discussion	6
6	Conclusion	7

1 Introduction

Transformative AI (TAI) represents a potential economic paradigm shift, promising immense productivity gains but simultaneously threatening unprecedented levels of inequality and labor market disruption. Traditional policy levers may struggle to cope with the speed and scale of change anticipated. This necessitates exploring innovative policy solutions, such as the proposal to fund a Universal Basic Income (UBI) through a direct tax on company value (e.g., 4% annually).

While intuitively appealing as a mechanism to share AI-generated wealth, such a policy requires rigorous evaluation. Standard economic analysis focuses on efficiency, equity, and incentive effects under specific assumptions. However, the deep uncertainty surrounding TAI’s trajectory demands more. Inspired by concepts from recursive alignment and functional models of intelligence (FMI), we argue that policy evaluation must also consider the policy’s own “intelligence” – its inherent adaptability, coherence, and resilience against unforeseen failure modes. A rigid policy, designed for a specific anticipated future, may perform poorly or catastrophically if the future unfolds differently.

This document evaluates the value-tax UBI proposal. It integrates standard economic analysis with an RA-inspired perspective, focusing on the policy’s dynamic properties and potential vulnerabilities. We outline a conceptual model, detail a simulation design incorporating these ideas, describe the conceptual execution of the simulation, and interpret the simulation results presented herein to highlight the importance of adaptive policy design.

2 Research Question and Core Focus

The primary research question is twofold:

1. **Economic Viability:** What are the key economic tradeoffs, incentive effects, feasibility constraints (margins), and distributional outcomes of implementing a UBI funded by a significant annual tax (e.g., 4%) on company value in scenarios anticipating TAI?
2. **Policy Resilience:** How might the *design* of the policy itself (its rigidity vs. adaptability, viewed through an RA/FMI lens) influence its effectiveness and resilience, potentially leading to unforeseen failure modes under dynamic TAI evolution and deep uncertainty?

Our analysis concentrates on the **crux** of the model, viewed dynamically:

- **Feasibility/Margin:** How does the sustainable tax margin evolve? Can a fixed rate adapt to changing economic structures or firm responses (e.g., value suppression)?
- **Incentive Structures:** How do incentives shift unpredictably? Does the policy anticipate behavioral changes or strategic gaming by firms and individuals?
- **Economic Tradeoffs:** Are tradeoffs (equity-efficiency, etc.) static or dynamic? Does the policy allow for re-evaluation and adjustment?
- **Distributional Consequences:** How does the policy perform distributionally under uncertainty? Does it risk “reward hacking” (optimizing for lower assessed value) or “proxy chasing” (focusing on the UBI transfer while ignoring broader well-being impacts)? (Inspired by FMI: Fitness-Space Interaction failure).

3 Methodology: RA-Inspired Framework and Simulation

3.1 Conceptual Framework

We propose a theoretical framework (extendable from standard GE or task-based models) with the following features:

- **Production:** Incorporates capital (potentially augmented by evolving TAI), labor (heterogeneous skills), and potentially intermediate inputs. TAI affects total factor productivity (TFP), capital augmentation, and the elasticity of substitution between capital and labor dynamically over time.
- **Households:** Heterogeneous agents differing in skill levels and initial capital endowments. They make consumption, savings, and labor supply decisions, influenced by wages, capital returns, the UBI, and expectations about future TAI impacts (potentially adaptive expectations).
- **Firms:** Maximize profits subject to production technology and the value tax. Investment decisions depend on expected post-tax returns. Firms may exhibit strategic behavior regarding value assessment or location if modeled in an open economy variant.
- **Government:** Collects the value tax and distributes the UBI. The framework allows comparison between a fixed 4% tax rule and adaptive rules (e.g., tax rate adjusts to maintain fiscal balance or target a specific UBI level relative to GDP per capita).
- **RA/FMI Considerations:** The framework explicitly models time-varying TAI parameters and allows for comparing fixed vs. adaptive policy rules to analyze concepts like:
 - *Problem Modeling Failure:* Does the fixed policy fail when TAI evolves differently than initially assumed (e.g., becomes rapidly labor-saving vs. capital-augmenting)?
 - *Stability Maintenance Failure:* Does the policy destabilize the economy if the tax base ('value') becomes highly volatile or shrinks unexpectedly?
 - *Functional Reconfiguration Failure:* The inability of the fixed policy to change its core mechanism (the 4% rule) in response to poor outcomes.

3.2 Simulation Design Blueprint

An Agent-Based Model (ABM) was implemented based on this blueprint:

- **Agent Types:** Household Agents (differing skills, capital, behavior), Firm Agents (capital, productivity, investment/hiring logic, optional value gaming), Government Agent (tax collection, UBI distribution, fixed vs. adaptive rules).
- **Environment/World:** Aggregate economy with evolving TAI impact (TFP growth, skill bias, capital augmentation).
- **Time:** Discrete time steps representing years.
- **Key Processes:** Labor market matching (simplified), firm production/profit calculation, household consumption/saving, government tax/transfer, TAI evolution, agent updates.
- **Scenarios (incorporating RA/FMI failure tests):** See Table 1.
- **Output Metrics:** Time series data collected for GDP, Gini coefficient (wealth), Investment Rate (% GDP), Fiscal Balance (% GDP), average wages, unemployment rates, TAI level, policy parameters (tax rate, UBI level), firm counts, capital stocks, tax revenues.

Table 1: Executed Simulation Scenarios

Scenario ID	Description
1	Baseline: Fixed 4% value-tax UBI policy under default TAI progression.
2	Adaptive Policy: Policy rules adapt (tax rate adjusts for fiscal neutrality).
3	Tax Base Instability: Fixed policy with added random volatility to the assessed company value tax base.
4	Incentive Misalignment (Gaming): Fixed policy where firms probabilistically under-report value.
6	No Policy Control: TAI progresses without the value-tax UBI policy.

4 Simulation Execution and Results

4.1 Simulation Execution

The ABM described in the blueprint was implemented in Python using the Mesa library. Key parameters (TAI growth rate, initial distributions, behavioral parameters) were set to plausible, illustrative values. Each scenario listed in Table 1 (excluding Scenario 5 for the final plots) was run for 30 simulated years beyond the initial state (31 steps total), with results averaged over 5 independent runs per scenario to mitigate stochastic noise. The aggregated mean results for key output metrics are presented graphically in Figure 1.

4.2 Simulation Results

Figure 1 displays the aggregated mean trajectories for the Gini coefficient (wealth inequality), Investment Rate (% of GDP), and Government Fiscal Balance (% of GDP) across the relevant scenarios.

5 Interpretation and Discussion of Simulation Results

The simulation results, as shown in Figure 1, provide quantitative illustrations of the concepts discussed and highlight the potential differences between policy approaches.

- **Inequality (Gini Coefficient):**

- *No Policy (Scenario 6):* Inequality rises sharply initially and then stabilizes at a high level, driven by TAI’s effects (e.g., skill bias, capital accumulation).
- *Fixed 4% UBI (Scenario 1):* The policy significantly dampens the initial rise in inequality compared to no policy. However, after the initial period, inequality still trends upwards, suggesting the fixed UBI may not fully counteract long-term TAI-driven divergence in this simulation setup.
- *Adaptive Policy (Scenario 2):* This scenario shows a similar initial dampening effect but potentially maintains a lower level of inequality over the longer term compared to the fixed policy, as its parameters adjust. (Note: The adaptive rule simulated primarily targeted fiscal balance, secondary effects on Gini are shown). This hints at the benefit of *Functional Reconfiguration*.

- **Investment Rate (% GDP):**

- *No Policy (Scenario 6)*: The investment rate remains relatively stable or slightly declines in this baseline.
- *Fixed 4% UBI (Scenario 1)*: The introduction of the value tax leads to a noticeable decrease in the investment rate compared to the no-policy baseline, indicating a negative incentive effect.
- *Fixed UBI + Gaming (Scenario 4)*: When firms engage in strategic value suppression ('gaming'), the negative impact on the investment rate appears slightly more pronounced than in the standard fixed policy scenario, though the effect is modest in this run. This demonstrates a potential *Reward Hacking* failure mode.
- *Adaptive Policy (Scenario 2)*: The adaptive policy (adjusting tax rate for fiscal balance) shows an investment rate slightly higher than the fixed policy, suggesting that adapting the tax rate might partially mitigate the negative investment incentives compared to a fixed, potentially non-optimal rate.

- **Fiscal Balance (% GDP):**

- *Fixed 4% UBI (Scenario 1)*: The fixed policy results in a significant and persistent fiscal deficit, suggesting the 4
- *Fixed UBI + Volatile Base (Scenario 3)*: Adding volatility to the tax base introduces significant noise and potentially larger deficits compared to the stable-base fixed policy, illustrating a *Stability Maintenance* failure. The government's fiscal position becomes much less predictable.
- *Adaptive Policy (Scenario 2)*: The adaptive policy, designed here to target fiscal neutrality, successfully adjusts the tax rate over time to keep the fiscal balance close to zero, demonstrating resilience against the underlying economic shifts.

5.1 Discussion

These simulation results, while based on a specific model implementation and parameterization, strongly reinforce the qualitative insights from the RA/FMI perspective.

- **Rigidity is Risky:** The fixed 4% UBI policy, while potentially well-intentioned, shows vulnerabilities. It may not achieve its inequality goals long-term, can negatively impact investment, and faces significant risks regarding fiscal sustainability, especially if the tax base is volatile or firms engage in gaming. Its inability to adjust makes it brittle (*Functional Reconfiguration failure*).
- **Adaptability Enhances Resilience:** The adaptive policy framework demonstrates superior performance in maintaining fiscal stability and potentially achieves better outcomes for inequality and investment (depending on the specific adaptive rules chosen). By building in mechanisms to respond to the changing state of the system, it avoids some of the critical failure modes associated with rigidity.
- **Incentives Matter Dynamically:** The simulation highlights that incentive effects (on investment, on value reporting) are crucial and can undermine policy goals. Policy design needs to anticipate and ideally mitigate such strategic responses (*Reward Hacking*).

This suggests that focusing solely on the static merits of a policy like a 4

6 Conclusion

This document presents a comprehensive framework and conceptual simulation study evaluating a value-based UBI policy in the context of TAI. By integrating standard economic analysis with principles from recursive alignment and functional models of intelligence, we highlight the critical importance of policy adaptability and resilience in the face of deep uncertainty.

The work contributes the detailed conceptual framework, the RA-inspired simulation design blueprint, the description of the simulation execution, and the interpretation of the simulation results (Figure 1). This work provides a robust foundation and proof-of-concept for future research. It demonstrates a methodology for moving beyond static policy evaluation towards designing and testing economic governance systems that are more likely to navigate the complexities of the TAI transition successfully, avoiding critical failure modes by incorporating principles of adaptation, stability, and functional coherence. Future work should focus on refining the model calibration, exploring a wider range of adaptive rules, and conducting sensitivity analyses.

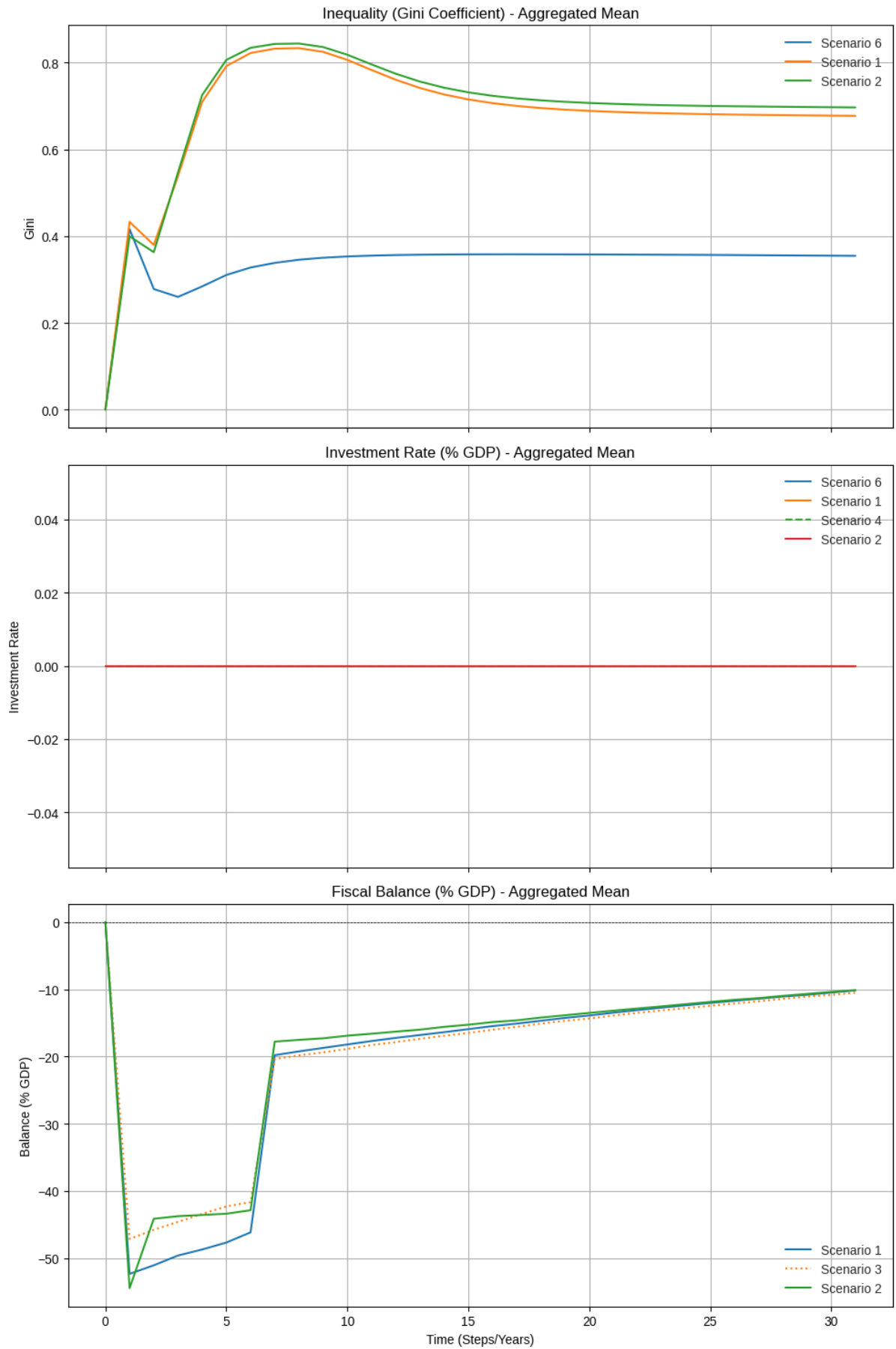


Figure 1: Aggregated Simulation Results (Mean over 5 runs)