AI Taxing AI: A Multi-Agent Reinforcement Learning Approach to Optimal Taxation

Peiyuan Zhang[a], Bo Li[a,*]

[a] *School of Economics, Peking University, Beijing, China*

Abstract

Artificial intelligence (AI) is improving productivity, yet its economy-wide diffusion may reshape the labor market and the distribution of income. To address these issues, we develop a Multi-Agent Reinforcement Learning (MARL) model featuring endogenous skill accumulation. We find that while AI boosts aggregate output via rapid process of skill sharing, it displaces middle-skilled humans and increases inequality. However, optimal taxation can redistribute the technological surplus to achieve Pareto improvements. To find the optimal tax schedule, we employ Dual-Loop Multi-Agent Reinforcement Learning to account for both government's fiscal policy and agent behaviors. Furthermore, we demonstrate that a "Digital Commons" regime—combining universal access with real-time knowledge sharing—maximizes social welfare by mitigating the trade-off between efficiency and equity.

Keywords: Artificial Intelligence, Optimal Taxation, Multi-Agent Reinforcement Learning, Inequality, Knowledge Diffusion

1. Introduction

Artificial intelligence (AI) is enhancing productivity in many domains, but its economy-wide diffusion also changes the policy problem faced by governments. How should fiscal policy respond when a new form of capital can replicate cognitive tasks and propagate knowledge rapidly? Can taxation transform the technological surplus into broad-based welfare gains rather than labor-market polarization? And how do knowledge-sharing rules and ownership structures shape the efficiency–equity trade-off?

*Corresponding author

Email address: libo@econ.pku.edu.cn (Bo Li)

The rapid proliferation of large language models (LLMs) and generative AI—exemplified by technologies such as GPT-4, Claude, and Gemini—suggests that AI is evolving into a general-purpose technology (GPT) (Brynjolfsson et al., 2017; Agrawal et al., 2019). Unlike earlier waves of automation that mainly substituted routine manual tasks, this wave of "cognitive automation" can perform non-routine analytical and creative tasks. Consequently, AI may reshape the task content of jobs, sectoral specialization, and the distribution of income. Recent evidence points to displacement pressures on middle-skilled (and some high-skilled) occupations, consistent with a potential "hollowing out" of the middle class (Acemoglu and Restrepo, 2020, 2022; **?**).

Furthermore, while optimal taxation can effectively mitigate inequality, it primarily addresses the *distributional* consequences of AI. It does not resolve the fundamental issue of *knowledge diffusion efficiency*. The current generative AI landscape is characterized by a stark dichotomy: proprietary, closed-source models developed by tech giants (e.g., OpenAI, Google) versus open-source, community-driven models (e.g., Meta's Llama, Mistral). This dichotomy introduces a critical trade-off. Closed-source models may incentivize high capital investment and computational infrastructure maintenance, but they create information silos that hinder economy-wide diffusion. Conversely, open-source models maximize the "velocity of knowledge" but face the risk of under-provision if maintenance incentives collapse. Therefore, addressing the economic impact of AI requires a framework that not only captures the dynamic interplay of labor supply and government redistribution but also evaluates how different ownership regimes and knowledge-sharing protocols shape aggregate welfare.

Designing policy in such an environment is challenging. Empirical evidence is inherently retrospective and offers limited counterfactual guidance for novel technologies and institutional designs. Theoretical approaches to optimal taxation and mechanism design, in turn, are constrained by analytical tractability and quickly become computationally infeasible once we incorporate rich heterogeneity and endogenous learning(Dyrda and Pedroni, 2023). These limitations motivate a machine-learning-based simulation approach that can generate counterfactuals and endogenize behavioral responses to policy.

To address these challenges, we develop a multi-agent reinforcement learning (MARL) framework to study optimal taxation in a model with endogenous skill accumulation and explicit knowledge diffusion. Building on the "AI Economist" paradigm (Zheng et al., 2022, 2021), we implement a dual-loop learning problem in which a government (social planner) chooses a state-contingent tax schedule while human and AI agents endogenously adjust labor supply and sectoral choices. This approach provides a tractable computational complement to theory for evaluating counterfactual tax policies when rich heterogeneity and non-linear transition dynamics make standard solution methods infeasible.

Our model embeds the economy in a spatial environment to capture transaction costs and mobility frictions, allowing equilibrium reallocation and transition dynamics to emerge from micro-level interactions rather than being imposed by reduced-form adjustment rules. The key economic mechanism is an asymmetry in learning and diffusion: human skills accumulate slowly through individual experience, whereas AI capabilities can be replicated and diffused across units at low marginal cost. This wedge governs the pace of technological diffusion, task reallocation, and the resulting distributional consequences.

We contribute to the literature along three dimensions. (i) *Mechanism:* we model an explicit learning-and-diffusion asymmetry between human and AI agents, which endogenously generates displacement and inequality dynamics. (ii) *Method:* we propose a dual-loop deep reinforcement learning approach that yields tractable solutions for state-contingent non-linear tax design in a high-dimensional dynamic economy. (iii) *Institutional design:* we compare knowledge-sharing and ownership regimes and quantify the welfare gains from treating general-purpose AI as a "Digital Commons".

1.1. Related Literature

Our work contributes to three strands of literature: (i) AI and labor market polarization, (ii) the optimal taxation of automation and AI, and (iii) applications of multi-agent reinforcement learning (MARL) in macroeconomics.

AI and Labor Market Polarization. A growing body of literature investigates how automation and AI reshape labor demand. Acemoglu and Restrepo (2018) and Acemoglu and Restrepo (2019) provide a task-based framework where automation displaces labor while new tasks reinstate it. Unlike traditional capital, AI is viewed as a "super-labor" that substitutes specifically for routine and increasingly cognitive tasks (Autor, 2013; Goos et al., 2014). Frey and Osborne (2017) estimate that a significant portion of jobs is susceptible to computerization. Our model extends this by endogenizing the skill accumulation process, explicitly differentiating between human "learning-by-doing" and AI's "collective intelligence" to simulate the dynamic substitution process.

Optimal Taxation of Automation. The debate on taxing robots to mitigate inequality is central to our policy analysis. Guerreiro et al. (2022) suggest that if robots are close substitutes for routine labor, taxing them can improve welfare by curbing wage inequality. **?** argue for a modest tax on robots effectively to manage the distributional consequences of new technologies. However, Thuemmel (2023) finds that the optimal robot tax might be zero or negative if robots complement high-skilled labor significantly. Our findings of a U-shaped tax schedule contribute to this debate by highlighting the non-linear optimal response to AI-induced displacement in a general equilibrium setting.

MARL in Economics. Methodologically, we follow the "AI Economist" framework proposed by Zheng et al. (2021, 2022), which demonstrates that two-level deep reinforcement learning can solve optimal taxation problems in dynamic environments with rich heterogeneity. Unlike standard heterogeneous agent models that rely on rational expectations, MARL allows for bounded rationality and complex spatial interactions (Foster et al., 2020). Similar approaches have been applied to auction design (**?**) and market design (Johanson et al., 2022), validating the efficacy of simulation-based economics.

The remainder of this paper is organized as follows. Section 2 details the economic environment, the distinct learning mechanisms of Human and AI agents, and Section 3 presents the dual-loop reinforcement learning framework. Section 4 presents the simulation results, analyzing the skill accumulation dynamics, the crowding-out effect, and the welfare implications of optimal taxation. Section 5 relaxes the baseline assumptions to explore the impact of ownership structures (private vs. public) and knowledge sharing protocols (open vs. closed) on economic outcomes. Finally, Section 6 concludes.

2. The Economic Model

We consider a dynamic spatial economy populated by N heterogeneous agents over a discrete time horizon $t = 0, 1, \dots, T$. The population $\mathcal{I}$ is partitioned into a set of human agents $\mathcal{H}$ and a set of AI agents $\mathcal{A}$, such that $\mathcal{I} = \mathcal{H} \cup \mathcal{A}$. The economy is governed by a Social Planner who implements tax and transfer policies to maximize social welfare.

2.1. Environment and State Space

The economy operates within a two-dimensional spatial grid $\mathcal{G}$ of size $L \times L$.

- **Resource Endowments:** The grid contains heterogeneous distributions of renewable natural resources (Wood and Stone). Agents born in different locations face varying initial endowments, naturally creating inequality in opportunity.
- **Spatial Frictions:** We introduce obstacles (represented as "Water") to model physical barriers. These obstacles increase the labor cost of movement, effectively simulating transportation costs.

Crucially, the terrain and resource distributions are **stochastically generated** in each episode to ensure that agents learn robust, generalizable economic policies rather than memorizing specific paths.

Agent State. At any time t, the state of an agent i is defined by the tuple $\Omega_{i,t} = \{\chi_{i,t}, \mathbf{x}_{i,t}, \mathbf{s}_{i,t}\}$, where:

- $\chi_{i,t} \in \mathcal{G}$: The agent's spatial coordinates (location).
- $\mathbf{x}_{i,t} = (x^w_{i,t}, x^s_{i,t}, x^c_{i,t})$: The inventory of physical resources (Wood, Stone) and liquid assets (Coin).
- $\mathbf{s}_{i,t} = (s^M_{i,t}, s^G_{i,t}, s^B_{i,t})$: The vector of accumulated skills for Move, Gather, and Build actions.

2.2. Technology and Actions

In each period t, agents allocate their time budget to a set of discrete economic actions: $\mathcal{A} = \{Move, Gather, Build, Trade\}$. These actions constitute a complete production-consumption cycle. The total labor supply $l_{i,t}$ is the sum of labor allocated to these specific tasks.

Search and Commute (Move). Agents navigate the spatial grid by moving in four cardinal directions (up, down, left, right). Movement is constrained by world boundaries and impassable terrain (Water). A successful move updates the agent's spatial state $\chi_{i,t}$. Economically, this mechanism explicitly micro-founds **search frictions** in the labor market: agents must physically expend resources to locate productive sites or trading venues. Crucially, we assume labor efficiency is endogenous to skill. The labor cost l^M decreases as the agent's movement skill s^M increases:

$$l^M(s^M_{i,t}) = l^M_0 \cdot \left(1 - \min\left(0.5, \frac{1}{1 - e^{-(s^M_{i,t} - \kappa_M)}}\right)\right) \tag{1}$$

This implies that highly skilled agents face lower effective search frictions, allowing them to traverse the market more efficiently.

Upstream Extraction (Gather). This action represents the **primary sector**. Upon locating a resource cell, agents supply labor l^G to extract raw materials. A successful gather action increases the agent's inventory $\mathbf{x}_{i,t}$ (Wood or Stone) and depletes the local resource stock. The technology is stochastic, with the success probability increasing in the agent's gathering skill s^G:

$$\text{Prob}(\Delta x^k_{i,t} = 1) = \min\left(1, \frac{1}{1 - e^{-(s^G_{i,t} - \kappa_G)}}\right), \quad k \in \{w, s\} \tag{2}$$

Here, skill accumulation acts as a technology shock that enhances the marginal productivity of labor in the extraction sector.

Downstream Manufacturing (Build). This action constitutes the **value-added manufacturing sector**. Agents combine intermediate inputs (1 unit of Wood, 1 unit of Stone) from their inventory with manufacturing labor l^B to produce the final good (Housing). The output generates income $y_{i,t}$, which is proportional to the agent's building skill s^B:

$$y_{i,t} = \begin{cases} \min(y_{max}, s^B_{i,t}) & \text{if } x^w_{i,t} \geq 1, x^s_{i,t} \geq 1 \text{ and } l^B_{i,t} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

This function captures the returns to high-skilled human capital: while raw materials are homogeneous, the value added by manufacturing depends critically on the agent's expertise.

Market Interaction (Trade). To reallocate intermediate goods according to comparative advantage, agents participate in a **centralized double-auction market**. Agents submit limit orders $\omega = \{type, quantity, price\}$ to a clearinghouse. The market mechanism automatically matches compatible bids and asks, facilitating resource exchange without requiring a bilateral coincidence of wants.

Sectoral Interdependence. The economy thus features a clear two-stage production structure. The upstream sector (Gatherers) relies on spatial mobility to extract dispersed natural resources, while the downstream sector (Builders) relies on acquiring these inputs to produce high-value final goods. The trade mechanism links these two sectors, allowing agents to endogenously specialize based on their evolving comparative advantages in friction reduction (Move/Gather skills) or value creation (Build skills).

2.3. Asymmetric Skill Accumulation

A central feature of our model is the structural asymmetry in human capital accumulation between biological humans and digital AI agents. This specification captures the fundamental difference in the learning mechanisms.

Human Agents: Path-Dependent Learning. We model human skill accumulation based on the *Learning-by-Doing* framework (Arrow, 1962) and the Human Capital theory of **?**. Human learning is characterized by high friction and path dependence. Knowledge is sticky and confined to the individual biological brain; a carpenter must practice for years to master their craft, and this skill cannot be simply "copied" to another person.

$$\mathbf{s}_{i,t} = \mathbf{s}_{i,t-1} + \rho \cdot \mathbf{l}_{i,t}, \quad \forall i \in \mathcal{H} \tag{4}$$

This linear accumulation reflects the reality of stratified education and experience-based proficiency.

AI Agents: Global Synchronization. In contrast, we model AI agents as algorithmic entities capable of **Collective Intelligence**. This setup mimics the current ecosystem of Large Language Models (LLMs). For instance, when a foundation model (e.g., GPT-4 or Gemini) undergoes a version update, every application calling its API experiences an instantaneous, leapfrog improvement in capability.

$$\mathbf{s}_{j,t} = \max_{k \in \mathcal{A}} \{\mathbf{s}_{k,t-1}\} + \rho \cdot \mathbf{l}_{j,t}, \quad \forall j \in \mathcal{A} \tag{5}$$

This "write-once, run-everywhere" mechanism captures the zero-marginal-cost scalability of digital intelligence. While this assumption of frictionless sharing is strong, it allows us to isolate the impact of the "velocity of knowledge" on the labor market. While our model environment resembles a resource economy, the 'Build' action represents complex cognitive tasks where AI's advantage lies in the zero-marginal cost of skill replication.

2.4. Optimization Problems

The Agent's Problem. Agents aim to maximize their expected discounted lifetime utility. We assume agents are "hand-to-mouth" consumers. The instantaneous utility function takes a standard CRRA form augmented with linear labor disutility:

$$u(c_{i,t}, l_{i,t}) = \frac{(c_{i,t})^{1-\sigma} - 1}{1-\sigma} - \eta \cdot l_{i,t} \tag{6}$$

Agents face a budget constraint where consumption cannot exceed disposable income after taxes and transfers ($c_{i,t} \leq y_{i,t}^{\text{net}}$). The agent chooses a policy π_i mapping states to actions to solve:

$$\max_{\pi_i} \mathbb{E}\left[\sum_{t=0}^{T} \beta^t u(c_{i,t}, l_{i,t})\right] \tag{7}$$

The Planner's Problem. The Social Planner seeks to optimize a Social Welfare Function (SWF) that balances aggregate productivity and equality. The planner observes the global state and sets a tax schedule τ and a lump-sum transfer Tr_t to maximize:

$$\max_{\tau} \mathcal{W} = \mathbb{E}\left[\sum_{t=0}^{T} \mathcal{P}_t(\tau) \times \mathcal{E}_t(\tau)\right] \tag{8}$$

where aggregate productivity $\mathcal{P}_t = \sum_i y_{i,t}$ and equality $\mathcal{E}_t = 1 - \text{Gini}(\{y_{i,t}\}_{i \in \mathcal{H}})$. The planner updates the tax policy every T_{tax} periods, creating a Stackelberg game structure.

3. Computational Strategy

Finding the equilibrium in this economy presents a significant computational challenge. The interdependence between the agents' labor supply decisions and the planner's tax policy creates a non-stationary environment. Formally, this is a **bi-level optimization problem**, akin to a dynamic Stackelberg game where the Social Planner (leader) sets the tax schedule, and the Agents (followers) respond by adjusting their labor supply and skill accumulation strategies.

To solve for this Stackelberg equilibrium, we employ a **Dual-Loop Multi-Agent Reinforcement Learning** framework (illustrated in Fig. 1). This framework iteratively solves the sub-games of the private sector and the planner:

The Inner Loop: Private Sector Response. In the inner loop, the tax schedule τ is held fixed. This phase simulates the short-run equilibrium where households and firms take government policy as given. Agents observe the current state $s_{i,t}$ and the fixed tax policy, then optimize their policy networks π_i to maximize their expected discounted lifetime utility. Through repeated interactions with the environment, agents learn to navigate spatial frictions, accumulate skills, and trade in the market, effectively mapping out their "best response" function to the current tax regime.

The Outer Loop: Optimal Policy Design. The outer loop represents the planner's optimization problem. The planner observes the aggregate state variables (e.g., income distribution, total output) resulting from the agents' behaviors in the inner loop. However, unlike standard static models, the planner does not adjust taxes continuously. Instead, policy updates occur every T_{tax} periods. This interval allows the private sector sufficient time to react and stabilize, reducing the noise in the planner's learning signal. At the end of each interval, the planner updates its policy network π_p via gradient ascent to maximize the Social Welfare Function (SWF),

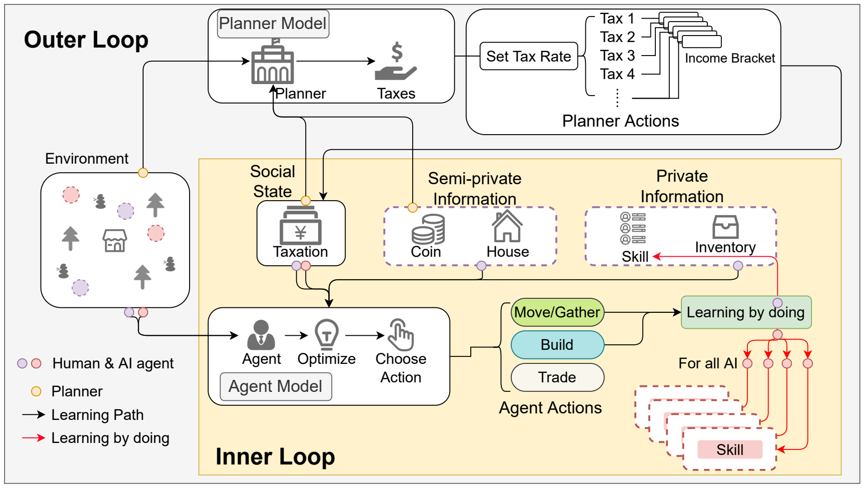


Figure 1: The Dual-Loop Learning Framework.

essentially performing a search over the space of possible tax functions to locate the global optimum.

3.1. Algorithm Implementation

We implement this framework using the *Proximal Policy Optimization* (PPO) algorithm (Schulman et al., 2017) within the RLlib library (Liang et al., 2018). PPO is chosen for its stability and sample efficiency, which are critical when training heterogeneous agents with continuous state spaces. The detailed execution flow is presented in Algorithm 1. By alternating between these two loops, the system gradually converges to a stationary point where agents play optimal strategies given taxes, and taxes are optimal given agent strategies.

3.2. Policy Network Architecture

To approximate the optimal policies for both agents and the planner, we employ deep neural networks capable of processing high-dimensional spatial and temporal data. Figure 2 illustrates the architecture used in our simulations.

The network is designed to handle the partial observability and spatial nature of the grid-world economy:

- **Spatial Feature Extraction (CNN):** The inputs include the map layer (terrain), resource layer (wood/stone), and agent layer (locations). We employ Convolutional Neural Networks (CNNs) to extract local spatial features, allowing agents to perceive their immediate surroundings and resource distributions.

Algorithm 1: Dual-Loop PPO for Optimal Taxation

Input: Episode Length T, Tax Frequency T_{tax}
Output: Planner's policy π_p, Agents' policy π

Initialize networks Φ_p (Planner), Φ (Agents);
for *Each Training Episode* **do**
    Reset World (s_0, o_0);
    Planner samples tax schedule $\tau \sim \pi_p(\cdot|s_0)$;
    for $t = 1$ **to** T **do**
        `// Inner Loop: Agents act`
        Agents choose actions $a_{i,t} \sim \pi(\cdot|s_{i,t}, \tau)$;
        Environment steps: $s_{t+1}, r_t \leftarrow \mathrm{Step}(a_t)$;
        `// Outer Loop: Planner updates`
        if $t \mod T_{tax} == 0$ **then**
            Planner applies tax and redistributes revenue;
            Planner updates policy $\tau' \sim \pi_p(\cdot|s_t)$ based on Social Welfare;
        end
        Store transitions for PPO training;
    end
    Update Φ and Φ_p using PPO gradients;
end

- **Temporal Processing (LSTM):** To capture the temporal dynamics of skill accumulation and market trends, the extracted features are fed into a Long Short-Term Memory (LSTM) layer. This recurrent structure enables agents to maintain a "memory" of past interactions, which is crucial for learning non-Markovian policies.

- **Decision Making (FFN):** Finally, Fully Connected (FC) layers map the hidden states to the output space. For agents, this results in a probability distribution over discrete actions (Move, Gather, Build, Trade); for the planner, it outputs a continuous vector defining the tax schedule.

3.3. Training Convergence and Stability

Before analyzing the economic implications, we verify the stability and convergence of our MARL framework. Since the agents and the planner are co-adapting in a non-stationary environment, ensuring that the system reaches a stable equilibrium is a prerequisite for valid economic interpretation.

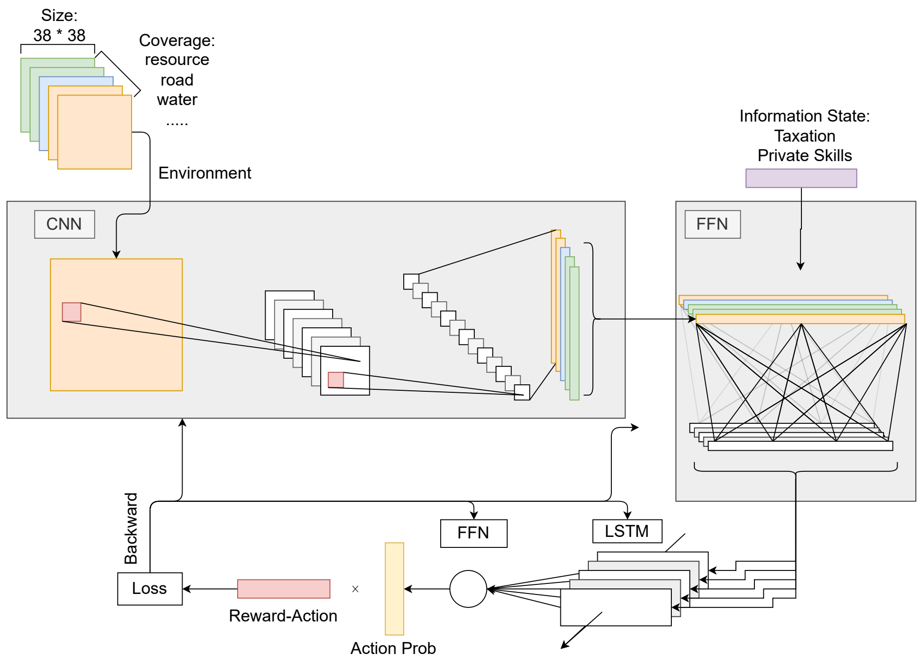


Figure 2: Schematic view of the Agent/Planner Policy Network.

Figure 3 displays the mean episode reward (social welfare) over training steps for the baseline and taxation experiments discussed in this section. We observe that:

1. **Rapid Initial Learning:** In the first phase (0-100k steps), agents rapidly learn basic survival skills (gathering and building), leading to a sharp increase in aggregate rewards.
2. **Policy Stabilization:** After approximately 300k steps, the reward curves for all scenarios (8H, 4H4A, etc.) plateau, indicating that both the private sector (agents) and the government (planner) have converged to their respective optimal policies.

Similarly, Figure 4 confirms the convergence for the ownership and knowledge-sharing experiments discussed in Section 5. The stability of these learning curves suggests that our reported results represent robust Stackelberg equilibria rather than transient training states.

4. Results: Inequality, Polarization, and Optimal Policy

The baseline model consists of eight human workers who learn optimal strategies through simulation. Simultaneously, the government sets the tax schedule to max-

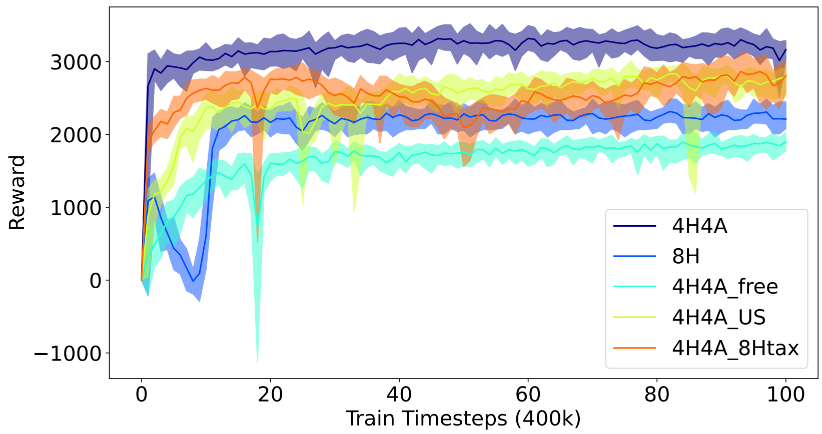


Figure 3: Empirical training progress for taxation models (Section 4). All models converge within 400k steps.

imize social welfare. To assess AI's impact on productivity and the labor market, and to evaluate potential tax policies, we conduct experiments across five distinct environments (Table 1) to isolate the effects of AI and taxation.

Table 1: Experimental Settings and Tax Regimes.

Model Name	Population Structure	Tax Policy Regime
8H (Baseline)	8 Human Agents	Optimal Taxation (Baseline)
4H4A	4 Humans + 4 AI	Optimal Taxation
4H4A_free	4 Humans + 4 AI	No Taxation
4H4A_US	4 Humans + 4 AI	US Federal Income Tax
4H4A_8HTax	4 Humans + 4 AI	Optimal Taxation (Baseline)

4.1. Asymmetric Skill Accumulation and Human Capital Stagnation

Figure 5 illustrates the divergence in skill accumulation, revealing the mechanism behind the "Great Divergence" between human capital and AI capabilities.

The AI Winner-Takes-All Dynamic. Subgraphs B1-B3 demonstrate that AI agents (dashed lines) achieve an exponential skill growth trajectory. This is a direct consequence of the collective intelligence mechanism: the marginal cost of skill acquisition for an individual AI agent is effectively zero, as it inherits the global frontier $\max_{m\in\mathcal{A}}\{s^k_{m,t-1}\}$ at the start of each period. This creates a "winner-takes-all" dynamic where AI capital rapidly saturates the high-skill sectors.

Figure 4: Empirical training progress for ownership models (Section 5).

The Rational Stagnation of Human Capital. Crucially, the impact on humans is heterogeneous, driven by the **return on skill investment**:

1. **High-Skilled Resilience (ID: 5):** Agents with initially high endowments maintain their skill accumulation. Their comparative advantage allows them to occupy niche markets where AI dominance is not yet absolute, preserving their incentive to invest in "learning-by-doing."
2. **Middle-Skilled Stagnation (ID: 7):** A key finding is the stagnation of the middle-income group. As AI rapidly substitutes for middle-skilled tasks (e.g., standard manufacturing), the market return to these skills diminishes. Facing reduced marginal returns on labor, these rational agents optimally reduce their labor supply, thereby halting their skill accumulation loop. Our simulation endogenously reproduces the 'hollowing out' phenomenon observed empirically: automation discourages human capital formation in sectors where humans lose comparative advantage.
3. **Low-Skilled Insularity:** Low-skilled agents remain largely unaffected in their accumulation trajectory, as they operate in low-value-added sectors (resource extraction) that AI agents, optimizing for high returns, largely ignore.

4.2. Endogenous Specialization

Fig. 6 decomposes the net income sources across three activities: Trade, Build, and Resource Collection. The top row (8H) depicts the economy without AI, while

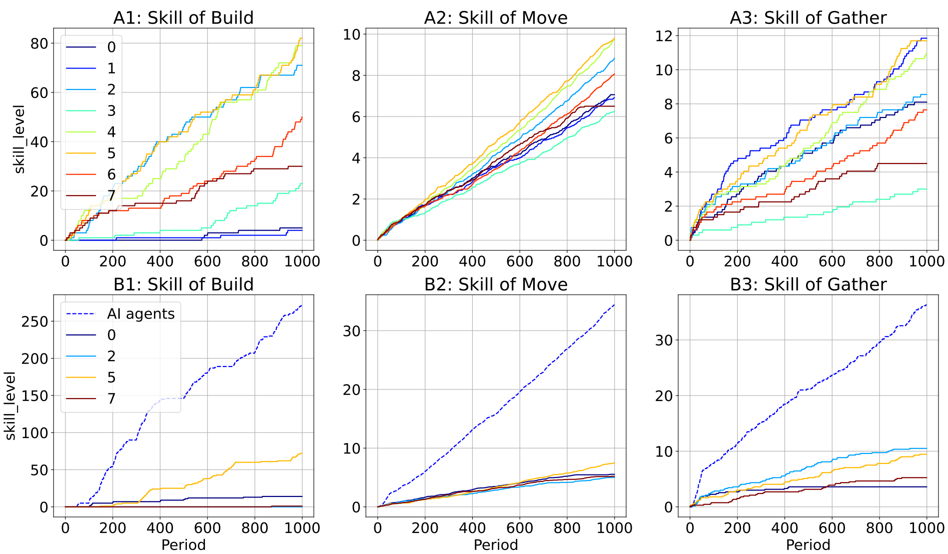


Figure 5: Skill Accumulation Dynamics.

the bottom row (4H4A) illustrates the AI-integrated economy across different tax regimes.

Baseline: Endogenous Specialization. In the human-only economy (Top Row), a clear complementary relationship emerges, grounded in comparative advantage. Low-skilled agents (left side of the x-axis) act as resource suppliers, generating positive trade income. In contrast, high-skilled agents (right side) experience negative trade income due to purchasing resources but achieve significant returns from Build. This pattern reflects a well-functioning supply chain, in which low-skilled agents provide raw materials while high-skilled agents specialize in high-value-added manufacturing.

AI Shock: Structural Crowding Out. The introduction of AI (bottom row) affects the labor market, as the clear distinction between human agents (circles) and AI agents (stars) reveals a pronounced segregation between the two groups:

- **Monopolization of Manufacturing:** As shown in Fig. 6 (row 2, Net income from Build), AI agents (stars) dominate the high-income region, whereas human agents (circles)—even those with relatively high skills—are suppressed to near-zero building income. AI's superior learning rate allows them to dominate the manufacturing sector.

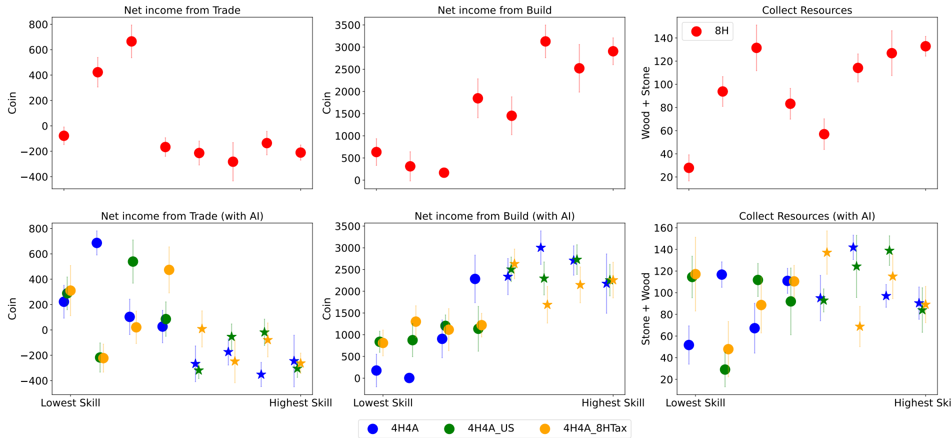


Figure 6: Income Source Decomposition by Skill Level (Top Row: 8H; Bottom Row: 4H4A).

- **Forced Retreat to Supply:** Consequently, humans are forced out of the value-added sector. The "Trade" panel shows that humans largely become net sellers, while AI agents become net buyers.

Mechanisms of Crowding Out: Substitution and Income Effects. The structural shift observed in Figure 6 is driven by two distinct economic forces:

- **The Substitution Channel (Comparative Advantage):** AI agents, with their superior learning rates, rapidly achieve a comparative advantage in the high-value-added "Building" sector. This lowers the relative price of manufactured goods and raises the opportunity cost for humans to engage in manufacturing. Consequently, humans are "crowded out" of the downstream sector and forced to retreat to the upstream "Trade/Gather" sector, acting as raw material suppliers to AI manufacturers.

- **The Income Effect (Leisure Preference):** The optimal tax policy (discussed below) extracts substantial surplus from AI and redistributes it to humans. For many middle-income agents, this non-labor income (transfer) increases their reservation wage. When the transfer payment creates a sufficient wealth effect, the marginal utility of consumption from additional labor falls below the marginal disutility of work (η). As a result, agents voluntarily withdraw from the labor market, effectively choosing leisure over low-return labor. This exacerbates the visible "crowding out" in labor supply statistics.

4.3. Income and Redistribution

Fig. 7 visualizes the wealth dynamics by comparing the baseline economy (A1-3, 8H) and the hybrid AI economy (B1-3, 4A4H).

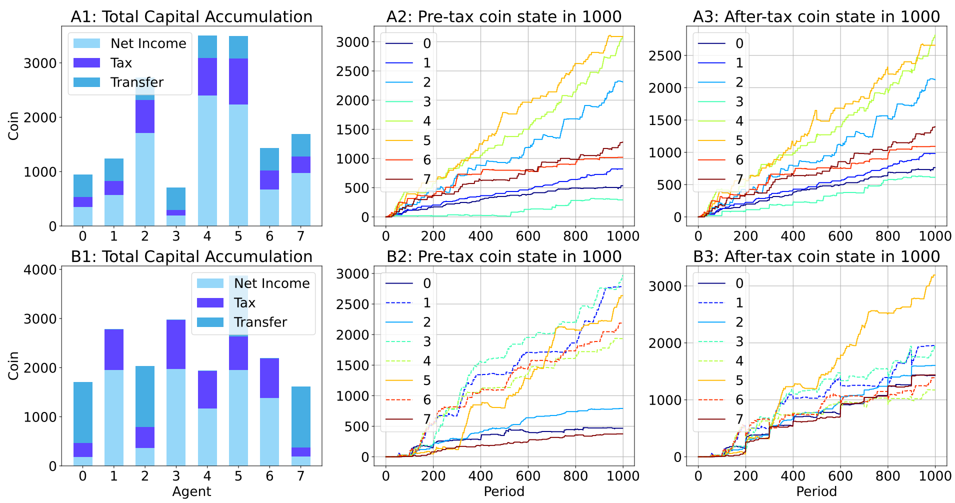
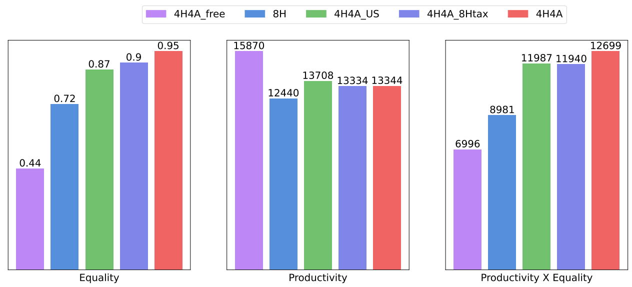


Figure 7: Wealth Accumulation Dynamics

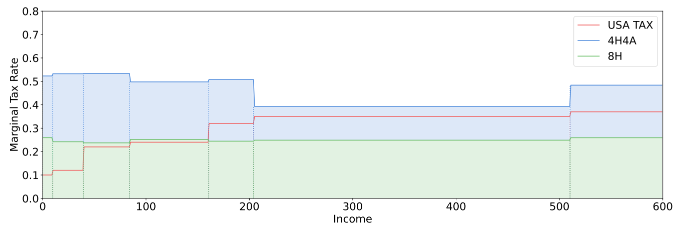
In the baseline economy (Fig. 7 - A2), wealth disparity is attributable solely to differences in spatial endowments. The introduction of AI (Fig. 7 - B2) induces a structural shift: while high-skilled agents (ID:5) remain resilient, the middle-income group (ID:7) faces severe displacement. Interestingly, low-skilled human agents experience a modest increase in their pre-tax income (Fig. 7 - B2). This occurs because low-skilled agents, remaining in resource extraction, benefit from selling materials to the AI manufacturing sector. Moreover, the planner can correct this distortion by implementing optimal taxation. (Fig. 7 - B3). By taxing high-yield AI agents, the planner redistributes the automated surplus, ensuring that even the displaced middle- and low-income agents achieve utility levels comparable to the baseline.

4.4. Welfare Analysis and Optimal Taxation

Does the integration of Artificial Intelligence merely redistribute existing value, or does it foster aggregate growth? Figure 8(a) (derived from the aggregate statistics in Table 1) provides quantitative evidence.



(a) Social Welfare Components



(b) Optimal Tax Schedules

Figure 8: Welfare and Policy Outcomes.

Aggregate Productivity Surge. Comparing the baseline human-only economy (8H) with the AI-integrated economy (4H4A), we observe a significant expansion in the production frontier. As shown in the "Productivity" bar chart, the total output increases from approximately 8,981 (8H) to 11,987 (4H4A), representing a **33.5% increase in aggregate productivity**. This confirms that AI acts as a potent productivity enhancer, effectively creating a "Technological Dividend."

Pareto Improvement via Redistribution. However, the "Equality" metric (Gini coefficient inverse) drops sharply from 0.72 to 0.44 in the absence of taxation (4H4A_free), reflecting extreme polarization. The core contribution of our optimal taxation framework is the restoration of equity without sacrificing efficiency. Under the optimal tax regime (4H4A), equality recovers to 0.95—surpassing even the human-only baseline—while productivity remains high (12,699). This demonstrates that the trade-off between efficiency and equity is not absolute; appropriate fiscal policy can convert the technological surplus into a Pareto improvement.

4.5. The Structure of Optimal Policy: Why a U-Shaped Tax?

Figure 8b illustrates the optimal marginal tax schedule (Blue line, 4H4A), which exhibits a distinct **U-shaped structure**, diverging significantly from the progressive steps of the standard US Federal Tax (Red line). This structure reveals an AI-discovered policy that aligns with the intuition of *The AI Economist* (Zheng et al., 2022) and classical Ramsey taxation principles:

- **High-Income Bracket (Taxing the Technological Rent):** The planner levies the highest marginal rates on top earners. In our model, these are exclusively AI agents. Economically, this targets the "quasi-rents" generated by the zero-cost collective intelligence mechanism. Since AI labor supply is driven by code rather than leisure preferences (subject to the maintenance constraints discussed in Section 5), their supply elasticity is relatively low, making this tax highly efficient and non-distortionary.
- **Middle-Income Bracket (Incentive Preservation):** The tax rate dips in the middle-income range. This "tax valley" is strategically designed to preserve the labor supply incentives for the remaining human workers. As we observed in the skill accumulation dynamics, middle-skilled humans are already facing severe substitution pressure from AI. A high marginal tax rate here would compound the negative effect, causing them to exit the labor market entirely (extensive margin response). By lowering the rate, the planner maintains their incentive to work.
- **Low-Income Bracket (Subsidization via Transfers):** While the nominal marginal rate appears high at the bottom, the system is anchored by a generous lump-sum transfer (funded by the high taxes on AI). This acts as a substantial subsidy (similar to a Negative Income Tax or UBI), guaranteeing that displaced workers maintain consumption levels comparable to the baseline.

In summary, the AI planner autonomously derives a tax scheme that aggressively redistributes the specific surplus generated by automation while carefully calibrating incentives to prevent the total collapse of human labor participation.

5. Open source and ownership of AI

While optimal taxation can effectively address the *distributional* consequences of AI, it treats the underlying industrial organization of the AI sector as exogenous. However, the current wave of generative AI reveals a complex landscape of ownership and access models—ranging from proprietary, closed-source models (like GPT-4) to open-weights community models (like Llama or Mistral).

This dichotomy presents a fundamental strategic trade-off. While proprietary models may secure the appropriability of high capital investments in compute and maintenance, they risk creating information silos that impede knowledge spillovers. Conversely, open-source architectures maximize the 'velocity of knowledge' yet remain susceptible to the tragedy of the commons should sustainable incentive structures for maintenance fail to materialize.

Therefore, merely taxing the output is insufficient; we must also ask: What is the optimal governance structure for this General Purpose Technology? In this section, we extend our framework to explore how property rights (Private vs. Public) and knowledge-sharing protocols (Open vs. Closed) influence the aggregate welfare of the economy.

AI's economic impact depends not only on the technology itself but also on its governing institutional frameworks. We now turn to the industrial organization of the AI sector: how do property rights and incentive structures influence the economic outcomes of AI? We extend our analysis along three critical dimensions: the *incentive compatibility* required to sustain AI productivity, the *velocity of knowledge diffusion* (Open vs. Closed), and the optimal *industrial organization* (Ownership Structure).

5.1. Endogenous Incentive Compatibility: The Optimal "Maintenance Share"

In the baseline analysis of Section 4, we implicitly assumed that AI capital is self-sustaining. This allowed us to isolate the effects of taxation on labor displacement. However, in reality, large-scale AI models require substantial continuous investment in computation and energy to maintain capability.

Typically, the decision of how much revenue to reinvest (ξ, the "Maintenance Share") is made by private firms maximizing profit. However, private profit maximization does not necessarily align with social welfare, especially when AI acts as a General Purpose Technology with broad distributional consequences. A private firm might set ξ too high (hoarding surplus for growth) or too low (extracting short-term dividends), ignoring the societal need for redistribution.

Therefore, in this part, we study the socially optimal maintenance share. We treat ξ not as a firm's control variable, but as a regulatory parameter determined by the Social Planner. The planner solves a Ramsey-type problem: finding the optimal ξ^* that balances the incentive constraint (keeping AI productive) against the redistributive motive (funding transfers to displaced humans).

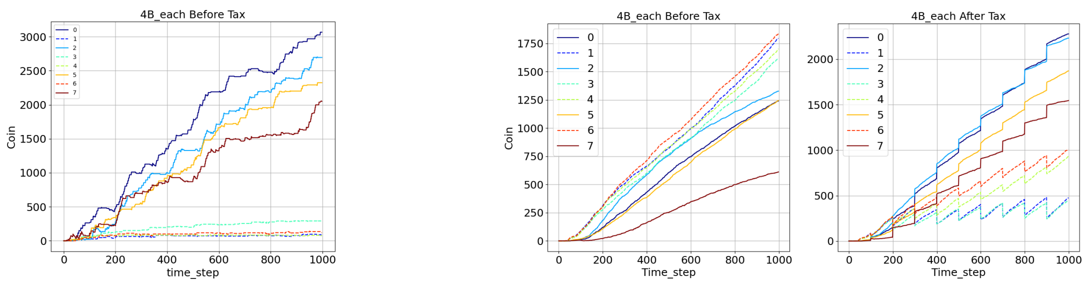
$$\max_{\tau,\xi} \mathcal{W} = \mathbb{E}\left[\sum_{t=0}^{T} \mathcal{P}_t(\tau,\xi) \times \mathcal{E}_t(\tau,\xi)\right] \tag{9}$$

This optimization problem reveals a fundamental trade-off:

- **Participation Constraint ($\xi < \xi^*$):** If the Maintenance Share is set too low (approaching 0), the incentive for AI agents to supply labor (compute) collapses. As shown in Fig. 9a, AI agents effectively withdraw from production, leading to an "Incentive Collapse" and a stagnation of aggregate output.

- **Redistributive Capacity ($\xi > \xi^*$):** Conversely, setting ξ too high reduces the fiscal capacity available for transfers. While AI productivity is maximized, the social planner has fewer resources to compensate displaced humans, thereby reducing aggregate social welfare (due to inequality aversion).

Therefore, the results presented in Fig. 9b reflect the economy under the **socially optimal Maintenance Share** ξ^*. Our simulations indicate that a strictly positive ξ^* is required to sustain the high productivity growth observed in our baseline while maximizing the funds available for redistribution.



(a) Incentive Collapse ($\xi \to 0$)

(b) Optimal Equilibrium ($\xi = \xi^*$)

Figure 9: Impact of the Maintenance Share (ξ) on AI Productivity. Panel (a) shows the collapse of production when incentives are insufficient; Panel (b) illustrates the sustained growth trajectory under the socially optimal Maintenance Share ξ^* determined by the planner.

5.2. The "Open vs. Closed" Debate: Velocity of Knowledge Diffusion

Conditional on a sustainable Maintenance Share (fixed at $\bar{\xi}$), we next address the debate between closed-source silos and open-source innovation. We fix the ownership structure to a distributed private model and vary the knowledge-sharing protocols across three specific scenarios:

1. **Private Silos:** AI agents only learn from their own experience, representing a completely closed-source ecosystem (4B_Ns).
2. **Periodic Sharing:** AI agents synchronize parameters only at fixed intervals (every T_{tax} periods), simulating a delayed open-source release cycle (4B_Ts).

Table 2: Experimental Scenarios and Economic Outcomes: Knowledge Sharing and Ownership Structures.

Scenario	Sharing Regime	Description	Aggregate Indicators		
			Welfare	Productivity	Equality
Panel A: Knowledge Sharing (Fixed Distributed Ownership)					
4B_Ns	**Private Silos**	No Sharing	8875.98	9628.27	0.91
4B_Ts	**Periodic Sharing**	Every T_{tax} periods	9267.07	9667.19	**0.96**
4B_Rs	**Digital Commons**	Real-time Sharing	9602.37	10749.49	0.89
Panel B: Ownership Structure (Fixed Real-time Sharing)					
4B_only	**Monopoly**	1 Human owns 4 AIs	9597.95	10318.31	0.93
2B2S	**Mixed Economy**	2 Private + 2 Public	10609.92	11616.45	0.91
4S	**Public Provision**	Fully Public	**11153.10**	**12040.80**	0.93

3. **Real-time Sharing (Digital Commons):** AI agents instantaneously share parameters globally, representing a fully open, real-time collaborative ecosystem (4B_Rs).

Results: The Velocity Premium. The results (Table 2 Panel A) demonstrate a clear hierarchy driven by information velocity. First, **Sharing outperforms Isolation**: both sharing regimes significantly outperform the private silos case in terms of productivity and welfare. Second, **Real-time Sharing outperforms Periodic Sharing**: the Real-time Sharing scenario yields higher aggregate output and welfare than Periodic Sharing. This confirms that in an AI-driven economy, the "velocity" of knowledge diffusion is a decisive factor. The "Digital Commons" mode, by minimizing information friction, maximizes the collective intelligence of the economy, providing strong support for open-source development paths.

5.3. Ownership Architecture: From Monopoly to Public Provision

Next, we investigate the impact of ownership. We compare four distinct ownership structures ranging from extreme privatization to full socialization:

1. **Monopoly:** Extreme concentration where one human agent owns all AI capital (4B_only).
2. **Distributed Private:** Decentralized ownership where every human owns one AI (4B_Rs).

3. **Mixed Economy:** A hybrid system with both private and public AI agents (2B2S).
4. **Public Provision:** All AI agents are publicly owned by the social planner (4S).

Results: Efficiency of the Digital Commons. Comparing these regimes reveals that the *Digital Commons* structure (4S) yields the optimal outcome in both aggregate productivity and equality (Table 2, Panel B).

Why does public provision strictly dominate proprietary ownership in this framework? The fundamental economic mechanism lies in the elimination of information frictions inherent in private models. In proprietary regimes (such as the Monopoly 4B_only or Distributed Private 4B_Rs), the misalignment between private profit maximization and social welfare leads to severe efficiency losses. Specifically, private owners are incentivized to restrict access, erect information silos, and under-invest in knowledge sharing to maintain their competitive advantage. This artificially limits the diffusion of AI capabilities, preventing the economy from fully exploiting the zero-marginal-cost scalability of digital intelligence.

In contrast, treating general-purpose AI as a *Digital Commons* allows the social planner to fully internalize the positive externalities of knowledge diffusion. By mandating universal access and real-time parameter synchronization, the public provision model maximizes the "velocity of knowledge." This structure eliminates the distortionary trade-offs found in ex-post private taxation models and ensures that the marginal benefit of AI deployment is equated across the entire economy. Consequently, this regime maximizes social welfare not merely through efficient redistribution, but by achieving first-best allocative efficiency in the ex-ante production process.

6. Conclusion

This paper studies optimal taxation and institutional design in an AI-driven economy where capabilities diffuse rapidly. We build a spatial multi-agent reinforcement learning (MARL) model with endogenous skill accumulation and explicit knowledge diffusion, and learn a state-contingent non-linear tax schedule jointly with agents' behavioral responses.

We obtain three main findings. First, an asymmetry in learning and diffusion—slow human learning-by-doing versus rapidly replicable AI capabilities—raises aggregate output but displaces middle-skilled workers and increases inequality. Second, optimal taxation can redistribute the technological surplus and deliver Pareto improvements relative to laissez-faire. Third, the efficiency–equity trade-off depends critically on information frictions: faster knowledge sharing increases welfare by raising the economy-wide "velocity of knowledge."

Institutional design is therefore first-order. Relative to private silos, knowledge sharing improves productivity and welfare, and real-time sharing dominates periodic synchronization. Sustaining openness, however, requires maintenance incentives: too little induces an incentive collapse, while too much erodes redistributive capacity. Finally, public provision with universal access and real-time sharing—a "Digital Commons"—internalizes diffusion externalities and outperforms proprietary ownership regimes.

Future work should incorporate endogenous innovation and market power in the AI sector, as well as political-economy constraints on implementation.

Declaration of Generative AI use

During the preparation of this work the author(s) used ChatGPT/Gemini in order to improve language and readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

References

Acemoglu, D., Restrepo, P., 2018. The race between man and machine: Implications of technology for growth, factor shares, and employment. American economic review 108, 1488–1542.

Acemoglu, D., Restrepo, P., 2019. Automation and new tasks: How technology displaces and reinstates labor. Journal of economic perspectives 33, 3–30.

Acemoglu, D., Restrepo, P., 2020. Robots and jobs: Evidence from us labor markets. Journal of Political Economy 128, 2188–2244.

Acemoglu, D., Restrepo, P., 2022. Tasks, automation, and the rise in us wage inequality. Econometrica 90, 1973–2016.

Agrawal, A., Gans, J., Goldfarb, A., 2019. The economics of artificial intelligence: An agenda. University of Chicago Press.

Arrow, K.J., 1962. The economic implications of learning by doing. The Review of Economic Studies 29, 155–173.

Autor, D.H., 2013. The “task approach” to labor markets: an overview. Journal for Labour Market Research 46, 185–199.

Brynjolfsson, E., Rock, D., Syverson, C., 2017. Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics. Working Paper 24001. National Bureau of Economic Research.

Dyrda, S., Pedroni, M., 2023. Optimal fiscal policy in a model with uninsurable idiosyncratic income risk. The Review of Economic Studies 90, 744–780.

Foster, I., Parkes, D., Zheng, S., 2020. The rise of ai-driven simulators: Building a new crystal ball. arXiv preprint arXiv:2012.06049 .

Frey, C.B., Osborne, M.A., 2017. The future of employment: How susceptible are jobs to computerisation? Technological forecasting and social change 114, 254–280.

Goos, M., Manning, A., Salomons, A., 2014. Explaining job polarization: Routine-biased technological change and offshoring. American economic review 104, 2509–2526.

Guerreiro, J., Rebelo, S., Teles, P., 2022. Should robots be taxed? The Review of Economic Studies 89, 279–311.

Johanson, M.B., Hughes, E., Timbers, F., Leibo, J.Z., 2022. Emergent bartering behaviour in multi-agent reinforcement learning. arXiv preprint arXiv:2205.06760 .

Liang, E., Liaw, R., Nishihara, R., Moritz, P., Fox, R., Goldberg, K., Gonzalez, J., Jordan, M., Stoica, I., 2018. Rllib: Abstractions for distributed reinforcement learning, in: International Conference on Machine Learning, PMLR. pp. 3053–3062.

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O., 2017. Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347 .

Thuemmel, U., 2023. Optimal taxation of robots. Journal of the European Economic Association 21, 1154–1190.

Zheng, S., Trott, A., Srinivasa, S., Parkes, D.C., Socher, R., 2022. The AI economist: Taxation policy design via two-level deep multiagent reinforcement learning. Science Advances 8, eabk2607.

Zheng, S., Trott, A., Srinivasan, S., Naik, N., Gruesbeck, M., Parkes, D.C., Socher, R., 2021. Building a foundation for data-driven, interpretable, and robust policy design using the AI Economist. arXiv preprint arXiv:2108.02904 .

Appendix A. Model Parameters and Hyperparameters

Table A.3 summarizes the key economic parameters and learning hyperparameters used in our multi-agent reinforcement learning (MARL) environment.

Table A.3: Summary of Baseline Model Parameters

Symbol	**Description**	**Value**
Demographics and Environment		
N	Total number of agents in the economy	8
$L \times L$	Spatial grid size	Random
T	Episode length (total time steps per episode)	1000
Preferences (Utility Function)		
β	Discount factor	0.99
σ	Coefficient of relative risk aversion (CRRA)	0.23
η	Marginal disutility of labor	0.21
Technology and Skill Accumulation		
l_0^M	Base labor cost for spatial movement	1.0
κ_M	Movement skill threshold/offset	10.0
κ_G	Gathering skill threshold/offset	10.0
y_{max}	Maximum output/income limit for the Build action	10.0
ρ	Skill accumulation rate (learning-by-doing increment)	0.1
Policy and AI Institutional Design		
T_{tax}	Frequency of planner policy updates (time steps)	100
$\bar{\xi}$	Fixed maintenance share for ownership analysis	0.6
MARL Training Hyperparameters (PPO)		
γ	PPO discount factor	0.98
lr	Learning rate for policy and value networks	0.0001
—	Hidden layer dimensions (FC)	[256, 256]