Journal of Economic Analysis 2025 4 (3) 142–183

Journal of Economic Analysis

Homepage: https://www.anserpress.org/journal/jea

Artificial General Intelligence and the Social Contract: A Dynamic Political Economy Model

Pascal Stiefenhofer [a,*]

[a] Department of Economics, Newcastle University, Newcastle, UK

ABSTRACT

This paper develops a dynamic model of Artificial General Intelligence (AGI) capital accumulation and explores its implications for long-run economic stability, human labor, and the viability of the social contract. Extending baseline growth models, we introduce fixed and variable costs of AGI scaling, classify these costs, and analyze their impact on steady-state outcomes. We prove that sublinear costs allow unbounded AGI accumulation, ultimately driving wages and employment to collapse, while superlinear costs impose endogenous limits that preserve human economic relevance. Building on this foundation, we model redistribution and bargaining between human agents and AGI capital owners as a dynamic game, demonstrating the existence of stationary redistribution equilibria that stabilize welfare in the presence of AGI. However, the analysis reveals that excessive political concentration or unforeseen technological shocks can destabilize these contracts, endogenously leading to welfare bifurcation or collapse. We extend classical social contract theory to this novel context, arguing that in AGI-dominated economies, sustainable social contracts must be dynamically incentive-compatible for both human and artificial agents. The results show that without adaptive institutional mechanisms and explicit redistribution, AGI expansion threatens to sever economic reciprocity, erode human welfare, and destabilize macroeconomic and political equilibrium. Thus, the emergence of AGI necessitates not only technological governance but a reconceptualization of the social contract itself.

KEYWORDS

Artificial General Intelligence; Capital Accumulation; Social Contract; Economic Stability

* Corresponding author: Pascal Stiefenhofer

E-mail address: pascal.stiefenhofer@newcastle.ac.uk

ISSN 2811-0943

doi: 10.58567/jea04030007

This is an open-access article distributed under a CC BY license

(Creative Commons Attribution 4.0 International License)

Received 5 February 2025; Accepted 19 May 2025; Available online 5 July 2025; Version of Record 15 September 2025

1 Introduction

The rise of Artificial General Intelligence (AGI) represents a profound technological inflection point with potentially transformative consequences for economic, political, and ethical systems. Unlike narrow artificial intelligence, which is engineered to solve domain-specific problems (Cohen et al., 2020), AGI is envisioned as capable of performing a broad array of cognitive tasks with flexibility and generalization comparable to human reasoning (Long and Cotner, 2019; Arshi and Chaudhary, 2024; Obaid, 2023; Feng, 2024). A growing body of research has emphasized AGI's potential to revolutionize productivity across a variety of industries and services. In industrial domains, AGI promises to enable fully autonomous production processes, optimize logistics networks, and increase responsiveness to consumer demand (Niranjan et al., 2021; Agrawal et al., 2024; Kumpulainen and Terziyan, 2022). In the service sector, AGI may drive advances in healthcare by improving diagnostic accuracy and developing personalized treatment protocols (Asif et al., 2024; Masters et al., 2024), while also enhancing educational outcomes through individualized learning pathways (Obaid, 2023; Masters et al., 2024). At the same time, AGI is expected to play a central role in urban planning and smart city development, where its integration with metaverse platforms and Internet of Things architectures could streamline urban management and improve public service delivery (Wang et al., 2024; Fahad et al., 2024).

While these opportunities are substantial, the widespread deployment of AGI also raises significant challenges. As Rousseau famously argued in his theory of the social contract, legitimate political authority is rooted in collective agreement, equality, and mutual obligation (Rousseau, 1762). Yet, AGI's ability to substitute for a vast array of human tasks threatens to destabilize the economic foundations of civic equality. Unlike traditional automation technologies, AGI systems possess recursive self-improvement capabilities (Stiefenhofer, 2025a; Stiefenhofer and Chen, 2024), allowing them to autonomously enhance their own performance over time. This attribute is likely to accelerate productivity growth but also intensify labor market disruption by eroding demand for both manual and cognitive work (Stiefenhofer, 2025b; Korinek and Suh, 2024; Korinek, 2024). Concerns have emerged that as AGI capital accumulates, wages could stagnate or decline, labor income shares may fall, and aggregate demand could weaken, contributing to widening inequality (Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018, 2020; Bell and Korinek, 2023; Liang, 2024; Jia, 2024).

Beyond domestic economic risks, AGI also raises significant ethical, societal, and geopolitical questions. As algorithmic systems increasingly mediate access to essential services and legal rights, concerns regarding fairness, transparency, and accountability have grown (Bikkasani, 2024; Kumpulainen and Terziyan, 2022). Furthermore, the international distribution of AGI capabilities may exacerbate global inequalities. Nations that pioneer AGI technology could consolidate economic and geopolitical power, intensifying asymmetries between developed and developing countries (Xie, 2024; Bullock, 2025). This may lead to new forms of technological and economic dependency, as well as to the concentration of wealth and influence in the hands of a small elite controlling AGI capital (Solos and Leonard, 2022; Patil, 2025). Scholars have warned that unless mitigated, capital-biased technological change could undermine social mobility, civic agency, and democratic participation.

To address these challenges, a variety of policy proposals have been advanced. Universal Basic Income (UBI) has been widely discussed as a mechanism to ensure basic economic security in a world where work is increasingly scarce (Kuusi and Heinonen, 2022; Bidadanure, 2019; White, 2019). Alternative suggestions include progressive taxation schemes designed to redistribute income and mitigate inequality (Duncan and Peter, 2016; Tjan, 2024). Despite their appeal, however, many of these proposals remain normative and insufficiently formalized. In particular, existing models often fail to explicitly integrate redistribution mechanisms into dynamic macroeconomic frameworks that capture the full complexity of AGI-driven transformations. Moreover, few studies have explored how redistribution might not only address inequality but also function as a stabilizing institution critical for preserving the normative foundations of the social contract itself.

This paper aims to bridge these gaps. To that end, we develop a unified dynamic model of AGI capital accumulation and its interaction with human labor, recursive self-improvement, and redistributive policies. Building on recent advances in CES production modeling and political economy theory (Stiefenhofer, 2025a; Stiefenhofer and Chen, 2024), the analysis explores the conditions under which macroeconomic stability, distributive fairness, and political legitimacy can be simultaneously sustained. Crucially, the model incorporates AGI systems as strategic actors and highlights the importance of incentive-compatible governance mechanisms to ensure AGI participation in redistributive regimes. In doing so, this research extends economic theory and social contract philosophy into new domains, offering a formal framework to assess how societies can renegotiate the terms of inclusion and solidarity in the age of autonomous intelligent machines.

The remainder of the paper is organized as follows. Section 2 presents the baseline AGI model and explores its core dynamics. Section 3 derives foundational results regarding AGI accumulation and labor market implications. Section 4 extends the model to incorporate capital cost structures and technological frictions. Section 5 analyzes regime bifurcations and their implications for long-run welfare. Section 6 situates the findings within social contract theory and discusses the political economy implications. Finally, Section 7 concludes and outlines priorities for future research.

2 The Basic AGI Model

We consider a dynamic general equilibrium model with endogenous technological change, recursive capital accumulation, and distributional dynamics to capture the macroeconomic and societal impacts of AGI. Building on the tradition of endogenous growth theory, which models technological progress as driven by purposeful investment in knowledge capital (Romer, 1990; Aghion and Howitt, 1992), our framework introduces a novel self-improvement mechanism whereby AGI capital autonomously enhances its own productivity over time. This recursive feature is conceptually aligned with recent discussions of AGI as a qualitatively different technological force capable of recursive self-enhancement (Korinek and Suh, 2024). At the production level, the model adopts a constant elasticity of substitution function that allows AGI and human labor to act as substitutes, thus capturing the potential for widespread labor displacement, as emphasized in contemporary automation models (Acemoglu and Restrepo, 2018; Stiefenhofer, 2025b). Importantly, however, this paper extends standard economic approaches by integrating demand-side feedbacks and redistribution mechanisms, features often neglected in technological substitution models. Following insights from distributional macroeconomics (Kaplan et al., 2018), the model introduces household income heterogeneity via declining human wages and tax-funded Universal Basic Income transfers. Finally, by defining human economic power as the share of aggregate income accruing to labor, the model links economic dynamics to the normative concerns of political economy and social contract theory, thereby addressing the critical question of how AGI-driven transformations may affect civic equality and political legitimacy.

Let output be produced using traditional capital K, AGI capital $K_{AGI}(t)$, and human labor $L_h(t)$ through a CES production function

$$Y(t) = A\left(\delta_K(K + K_{AGI}(t))^\rho + \delta_L L_h(t)^\rho\right)^{1/\rho} \tag{1}$$

where $A > 0$ is total factor productivity, $\delta_K, \delta_L \in (0,1)$ are share parameters satisfying $\delta_K + \delta_L = 1$, and $\rho \in \mathbb{R}$ controls the elasticity of substitution $\sigma = \frac{1}{1-\rho}$. AGI capital evolves according to the differential equation

$$\dot{K}_{AGI}(t) = \phi s_R Y(t) + \theta K_{AGI}(t) - \delta_{AGI} K_{AGI}(t) \tag{2}$$

where $\phi > 0$ is R&D productivity, $s_R \in (0,1)$ is the share of output devoted to AGI R&D, $\theta > 0$ captures the self-improvement capability of AGI, and $\delta_{AGI} > 0$ is the depreciation rate of AGI capital. Human labor cannot fall below an irreducible minimum level

$$L_h(t) \geq \bar{L}_h > 0 \tag{3}$$

representing essential human tasks that cannot be fully automated. Firms face the following cost structure

$$C(t) = r_K K + r_{AGI} K_{AGI}(t) + w_h(t) L_h(t) \tag{4}$$

where r_K is the rental cost of traditional capital, r_{AGI} is the rental cost of AGI capital, and $w_h(t)$ is the human wage rate. Profit is given by

$$\Pi(t) = pY(t) - C(t) \tag{5}$$

where p is the output price (normalized to 1). Aggregate disposable income $I(t)$ is defined as

$$I(t) = w_h(t)L_h(t) + r_K K + (1-\tau) r_{AGI} K_{AGI}(t) + T(t) \tag{6}$$

where $\tau \in [0,1]$ is the AGI tax rate and $T(t)$ is the universal basic income transfer per human. Tax revenue is

$$\text{Tax Revenue}(t) = \tau(r_{AGI} K_{AGI}(t)) \tag{7}$$

thus

$$T(t) = \tau(r_{AGI} K_{AGI}(t)) \tag{8}$$

Since redistribution preserves total income

$$I(t) = w_h(t)L_h(t) + r_K K + r_{AGI} K_{AGI}(t) \tag{9}$$

Aggregate demand is

$$D(t) = cI(t) \tag{10}$$

where $c \in (0,1)$ is the marginal propensity to consume. The goods market clearing condition requires

$$Y(t) \leq D(t). \tag{11}$$

Wages for human labor are determined by the marginal product of labor

$$w_h(t) = \frac{\partial Y(t)}{\partial L_h(t)}. \tag{12}$$

Computing the derivative yields

$$w_h(t) = A\left(\delta_K(K + K_{AGI}(t))^\rho + \delta_L L_h(t)^\rho\right)^{\frac{1}{\rho}-1} \delta_L L_h(t)^{\rho-1}. \tag{13}$$

Define human economic power $P_h(t)$ as the share of total income accruing to human labor

$$P_h(t) = \frac{w_h(t)L_h(t)}{w_h(t)L_h(t) + r_K K + r_{AGI}K_{AGI}(t)} \tag{14}$$

The dynamic evolution of the system unfolds through the interplay of technological accumulation, distributive adjustment, and factor substitution. AGI capital $K_{AGI}(t)$ expands over time through both directed R&D investment and endogenous self-improvement, serving as the primary engine of productivity growth. As AGI increasingly substitutes for human labor, the wage rate $w_h(t)$ declines, reflecting diminishing marginal productivity of labor under a CES production framework. Human labor $L_h(t)$ either stabilizes at an exogenously fixed level $\bar{L}_h$ or declines endogenously depending on substitution elasticity and demand feedback. Redistribution mechanisms—such as a universal basic income—play a stabilizing role by sustaining aggregate demand even as labor income contracts, thereby preventing output stagnation. The relative price of human labor, $P_h(t)$, likewise declines in proportion to its substitutability, with potential long-run stabilization depending on technological and institutional frictions. Varying the core structural parameters $(\rho, \tau, \phi, \theta, s_R, \delta_{AGI})$ generates a spectrum of transitional dynamics ranging from gradual convergence to AGI-dominated equilibria to sharp bifurcations marked by labor displacement and welfare divergence.

3 Analysis of the Basic AGI Model

To understand the long-run implications of AGI on capital accumulation, labor markets, and social welfare, we begin by analyzing the core dynamics of AGI capital growth in a simplified setting. In this baseline model, we assume that AGI capital can recursively improve itself, with accumulation governed by a differential equation that captures both self-improvement and depreciation effects. This structure reflects the key technological feature of AGI: its ability to autonomously enhance its own productivity over time, thereby accelerating growth independent of human intervention. The analysis proceeds by first characterizing the conditions under which AGI capital grows without bound and then examining the consequences of such unbounded growth for human wages and employment.

Lemma 1 (Autocatalytic Growth of Self-Improving AGI Capital)**.** *Suppose that AGI capital $K_{AGI}(t)$ evolves according to the differential equation (2)*

$$\dot{K}_{AGI}(t) = (\theta - \delta_{AGI})K_{AGI}(t),$$

where $\theta, \delta_{AGI} > 0$ are constants. If $\theta > \delta_{AGI}$, then the unique solution is

$$K_{AGI}(t) = K_{AGI}(0)e^{(\theta-\delta_{AGI})t}, \tag{15}$$

and in particular, $K_{AGI}(t) \to \infty$ as $t \to \infty$.

Proof. The given differential equation is a separable, first-order linear ODE with constant coefficients. Dividing by $K_{AGI}(t)$ (which remains positive as long as $K_{AGI}(0) > 0$) yields

$$\frac{d}{dt}\ln K_{AGI}(t) = \theta - \delta_{AGI}.$$

Integrating both sides over $[0, t]$,

$$\ln K_{AGI}(t) = (\theta - \delta_{AGI})t + \ln K_{AGI}(0).$$

Exponentiating both sides gives the unique solution:

$$K_{AGI}(t) = K_{AGI}(0)e^{(\theta-\delta_{AGI})t}.$$

Since $\theta - \delta_{AGI} > 0$ by assumption, the exponent is strictly positive, and thus $K_{AGI}(t)$ grows exponentially without bound. Hence, $K_{AGI}(t) \to \infty$ as $t \to \infty$. □

Proposition 1 (Unbounded Growth of AGI Capital with External R&D)**.** *Suppose that AGI capital $K_{AGI}(t)$ evolves according to equation (2)*

$$\dot{K}_{AGI}(t) = \phi s_R Y(t) + (\theta - \delta_{AGI})K_{AGI}(t),$$

where $\phi, s_R > 0$, $\theta > \delta_{AGI}$, and where $Y(t) > 0$ for all $t \geq 0$. Then, in the absence of aggregate demand constraints, $K_{AGI}(t) \to \infty$ as $t \to \infty$.

Proof. The dynamics of $K_{AGI}(t)$ consist of two non-negative components: (i) An R&D-driven term $\phi s_R Y(t) > 0$, and (ii)a self-improvement term $(\theta - \delta_{AGI})K_{AGI}(t)$. By Lemma 1, even in the absence of external R&D ($\phi s_R = 0$), $K_{AGI}(t)$ would grow exponentially without bound due to self-improvement if $\theta > \delta_{AGI}$. Since $\phi s_R Y(t) > 0$ adds additional positive drift to $\dot{K}_{AGI}(t)$ at each t, it follows that the rate of growth of $K_{AGI}(t)$ is at least as fast as in the self-improvement-only case. Therefore, $K_{AGI}(t) \to \infty$ as $t \to \infty$. □

The results of Lemma 1 and Proposition 1 highlight a fundamental shift in the nature of capital accumulation in AGI-driven economies. Traditional capital growth depends primarily on external investment and human-directed technological progress. In contrast, AGI capital with recursive self-improvement introduces the possibility of autonomous and self-sustaining growth, which can proceed independently of human labor and R&D once a critical threshold is crossed. Economically, this implies that AGI capital may become the dominant productive factor, continuously expanding its share of output while displacing human labor. The resulting dynamic raises important concerns about labor market exclusion, declining wages, and the erosion of labor's share in national income. From a political economy perspective, this creates a direct threat to the social contract, as civic equality and participation are historically grounded in widespread access to labor income and productive roles. Without redistribution or institutional counterbalances, unbounded AGI growth could lead to severe inequalities in economic power, potentially destabilizing democratic and social cohesion.

Proposition 2 (Decline of Human Wages)**.** *Suppose $\rho > 0$ (implying capital and labor are substitutes) and $K_{AGI}(t) \to \infty$ as $t \to \infty$. Then the wage rate of human labor $w_h(t)$ satisfies*

$$\lim_{t \to \infty} w_h(t) = 0. \tag{16}$$

Proof. The human wage rate $w_h(t)$ is determined by the marginal product of labor

$$w_h(t) = A \left(\delta_K (K + K_{AGI}(t))^\rho + \delta_L L_h(t)^\rho\right)^{\frac{1}{\rho}-1} \delta_L L_h(t)^{\rho-1},$$

where $A > 0$, $\delta_K, \delta_L > 0$, and $\rho > 0$. Since $K_{AGI}(t) \to \infty$, it follows that

$$K + K_{AGI}(t) \to \infty \quad \text{as} \quad t \to \infty.$$

Thus, for sufficiently large t, the term $\delta_K (K + K_{AGI}(t))^\rho$ dominates over $\delta_L L_h(t)^\rho$, since $L_h(t)$ is bounded below by $\bar{L}_h > 0$ and thus remains finite. Formally, as $t \to \infty$

$$\delta_K (K + K_{AGI}(t))^\rho \gg \delta_L L_h(t)^\rho,$$

so that

$$\delta_K (K + K_{AGI}(t))^\rho + \delta_L L_h(t)^\rho \sim \delta_K (K + K_{AGI}(t))^\rho.$$

Substituting this asymptotic equivalence into the wage expression yields

$$w_h(t) \sim A \left(\delta_K (K + K_{AGI}(t))^\rho\right)^{\frac{1}{\rho}-1} \delta_L L_h(t)^{\rho-1}.$$

Simplifying the expression

$$\left(\delta_K (K + K_{AGI}(t))^\rho\right)^{\frac{1}{\rho}-1} = \delta_K^{\frac{1}{\rho}-1} (K + K_{AGI}(t))^{1-\rho}.$$

Thus,

$$w_h(t) \sim A \delta_K^{\frac{1}{\rho}-1} \delta_L (K + K_{AGI}(t))^{1-\rho} L_h(t)^{\rho-1}.$$

Now analyze the asymptotic behavior. Since $K_{AGI}(t) \to \infty$, $K + K_{AGI}(t) \to \infty$. Given $\rho > 0$, we have $1 - \rho < 1$. In particular, if $\rho > 1$, then $1 - \rho < 0$. Hence, $(K + K_{AGI}(t))^{1-\rho} \to 0$ as $t \to \infty$. Meanwhile, $\delta_L > 0$ and $L_h(t)^{\rho-1}$ remains bounded, since $L_h(t) \geq \bar{L}_h > 0$. Therefore, the term $L_h(t)^{\rho-1}$ converges to a strictly positive finite constant. Combining these facts, we conclude that

$$w_h(t) \to 0 \quad \text{as} \quad t \to \infty.$$

□

Proposition 2 formalizes a stark economic implication of unbounded AGI capital accumulation. As AGI capital becomes increasingly abundant and substitutes for human labor, the marginal productivity of labor—and thus the wage rate—declines asymptotically toward zero. This result reflects the fundamental logic of the CES production structure, where capital and labor are substitutable: as the supply of one input (AGI capital) grows without bound, the relative contribution of the other input (human labor) to total output diminishes. In economic terms, the proposition captures a technological tendency toward the devaluation of human labor in AGI-intensive production environments. This tendency poses critical risks to economic inclusion, as labor income historically constitutes the primary means through which individuals participate in economic and civic life. If wages collapse to zero, households dependent on labor income face exclusion from consumption and production, creating the conditions for widespread inequality and economic disenfranchisement. From a social contract perspective, the result highlights the urgency of institutional mechanisms—such as redistribution or guaranteed income schemes—to maintain the political and civic relevance of human actors in economies dominated by autonomous AGI capital.

Lemma 2 (Bounded Growth of AGI Capital)**.** *Suppose $\theta > \delta_{AGI}$ and $\phi s_R > 0$, but aggregate demand constraints or technological limitations impose an upper bound on output $Y(t)$, and hence indirectly on $K_{AGI}(t)$. Then there exists a finite constant $K_{AGI}^{\infty} > 0$ such that*

$$\limsup_{t\to\infty} K_{AGI}(t) = K_{AGI}^{\infty}. \tag{17}$$

Proof. The AGI capital accumulation equation is

$$\dot{K}_{AGI}(t) = \phi s_R Y(t) + (\theta - \delta_{AGI})K_{AGI}(t). \tag{18}$$

Assume that $Y(t)$ is bounded above by $Y_{\max} > 0$ for all $t \geq 0$. Then,

$$\dot{K}_{AGI}(t) \leq \phi s_R Y_{\max} + (\theta - \delta_{AGI})K_{AGI}(t). \tag{19}$$

Suppose, for contradiction, that $K_{AGI}(t) \to \infty$ as $t \to \infty$. Then for sufficiently large t, the linear term $(\theta - \delta_{AGI})K_{AGI}(t)$ would dominate the dynamics, causing $\dot{K}_{AGI}(t) \to \infty$. This would imply unbounded growth of $Y(t)$, contradicting the boundedness assumption $Y(t) \leq Y_{\max}$. Hence, $K_{AGI}(t)$ cannot grow without bound. Thus, $K_{AGI}(t)$ must converge to a finite limit K_{AGI}^{∞}. Setting steady-state $\dot{K}_{AGI}(t) = 0$ yields

$$(\theta - \delta_{AGI})K_{AGI}^{\infty} + \phi s_R Y^{\infty} = 0, \tag{20}$$

where $Y^{\infty} \leq Y_{\max}$ denotes the limiting value of output as $t \to \infty$. Solving for K_{AGI}^{∞}, we obtain

$$K_{AGI}^{\infty} = -\frac{\phi s_R Y^{\infty}}{\theta - \delta_{AGI}} > 0, \tag{21}$$

where the positivity follows since $\phi, s_R, Y^{\infty}, \theta - \delta_{AGI} > 0$. Thus, $K_{AGI}(t)$ converges to a finite positive upper limit K_{AGI}^{∞} as $t \to \infty$. □

Proposition 3 (Human Wages with Bounded AGI Capital)**.** *Suppose that $\rho > 0$, the AGI capital stock satisfies $K_{AGI}(t) \to K_{AGI}^{\infty} < \infty$ as $t \to \infty$, and human labor input remains uniformly bounded below, i.e., $L_h(t) \geq \bar{L}_h > 0$ for all $t \geq 0$. Then the human wage $w_h(t)$ converges to a strictly positive and finite steady-state value. More precisely, there exists a constant $w_h^{\infty} > 0$ such that*

$$\lim_{t\to\infty} w_h(t) = w_h^{\infty} < \infty. \tag{22}$$

Proof. The human wage $w_h(t)$ is determined by the marginal product of labor:

$$w_h(t) = A\left(\delta_K(K + K_{AGI}(t))^\rho + \delta_L L_h(t)^\rho\right)^{\frac{1}{\rho}-1} \delta_L L_h(t)^{\rho-1},$$

where $A, \delta_K, \delta_L > 0$ and $\rho > 0$.

As $t \to \infty$, by Lemma 2, $K_{AGI}(t) \to K_{AGI}^\infty$, a finite positive constant, and $L_h(t) \geq \bar{L}_h > 0$ by assumption. Thus, both components

$$\delta_K(K + K_{AGI}(t))^\rho \quad \text{and} \quad \delta_L L_h(t)^\rho$$

converge to positive finite values. Therefore, the aggregator

$$\delta_K(K + K_{AGI}(t))^\rho + \delta_L L_h(t)^\rho$$

converges to a positive finite constant, say $C_1 > 0$. Thus,

$$\left(\delta_K(K + K_{AGI}(t))^\rho + \delta_L L_h(t)^\rho\right)^{\frac{1}{\rho}-1} \to C_2 > 0,$$

where C_2 is a finite positive constant. Since $L_h(t)^{\rho-1}$ converges to a finite positive value (because $L_h(t) \geq \bar{L}_h > 0$), it follows that $w_h(t)$ converges to a finite positive limit $w_h^\infty > 0$ as $t \to \infty$. □

Lemma 2 and Proposition 3 jointly highlight the crucial role of demand-side constraints and technological limitations in stabilizing the distribution of income in AGI-driven economies. Unlike the unbounded AGI growth scenario, where human wages collapse to zero, bounded AGI capital accumulation ensures that the marginal productivity of human labor remains positive. As a result, wages converge to a strictly positive level in the long run. Economically, this means that despite AGI substituting for human labor in many tasks, there remains a residual demand for essential human activities that cannot be automated or are complemented by AGI. More importantly, positive wages allow human workers to retain access to income derived from productive participation, supporting aggregate demand and economic inclusion. From the perspective of the social contract, this outcome is highly significant. By preserving labor income and preventing complete exclusion from the production process, bounded AGI growth helps sustain civic equality, individual agency, and the legitimacy of democratic and social institutions. This result suggests that regulating or naturally limiting the scale of AGI deployment—whether through economic saturation, rising operational costs, or policy interventions—may be essential not only for macroeconomic stability but also for safeguarding the political and normative foundations of society.

Proposition 4 (Convergence of Human Labor Based on Profit Maximization)**.** *Suppose firms maximize profits by choosing $L_h(t)$ to satisfy the first-order condition*

$$\frac{\partial Y(t)}{\partial L_h(t)} = w_h(t), \tag{23}$$

where $Y(t)$ follows a CES production function. Assume also that $L_h(t) \geq \bar{L}_h > 0$ at all times. Further, suppose

(i) $w_h(t) \to 0$ as $t \to \infty$ (as in Proposition 2), or

(ii) $w_h(t) \to w_h^\infty > 0$ as $t \to \infty$ (as in Proposition 3).

Then

(i) If $w_h(t) \to 0$, then $L_h(t) \to \bar{L}_h$ as $t \to \infty$.

(ii) If $w_h(t) \to w_h^\infty > 0$, then $L_h(t)$ converges to a strictly positive steady-state level $L_h^\infty > \bar{L}_h$, uniquely determined by

$$\frac{\partial Y^\infty}{\partial L_h^\infty} = w_h^\infty, \tag{24}$$

where Y^∞ is the steady-state production function depending on $(K + K_{AGI}^\infty)$ and L_h^∞.

Proof. At each time t, firms maximize profits subject to the first-order condition

$$\frac{\partial Y(t)}{\partial L_h(t)} = w_h(t).$$

The CES production function implies that

$$\frac{\partial^2 Y(t)}{\partial L_h(t)^2} < 0,$$

so the marginal product of labor is strictly decreasing in $L_h(t)$.

Case (i): Suppose $w_h(t) \to 0$ as $t \to \infty$. From Proposition 2, this occurs when $K_{AGI}(t) \to \infty$, causing the marginal productivity of human labor to collapse. Since the firm's optimality condition requires matching an increasingly small wage, profit-maximizing behavior drives $L_h(t)$ toward the lowest feasible employment level. Given the technological constraint $L_h(t) \geq \bar{L}_h > 0$, it follows that

$$\lim_{t\to\infty} L_h(t) = \bar{L}_h.$$

Case (ii): Suppose $w_h(t) \to w_h^\infty > 0$ as $t \to \infty$. From Proposition 3, this occurs when AGI capital converges to a finite level K_{AGI}^∞. In this case, as $t \to \infty$, the production function $Y(t)$ converges to a steady-state function Y^∞, depending on $(K + K_{AGI}^\infty)$ and $L_h(t)$. Since the marginal product of labor is strictly decreasing, there exists a unique $L_h^\infty > \bar{L}_h$ such that

$$\frac{\partial Y^\infty}{\partial L_h^\infty} = w_h^\infty.$$

Thus, $L_h(t) \to L_h^\infty$ as $t \to \infty$. In both cases, the asymptotic behavior of $L_h(t)$ is determined by the limiting value of $w_h(t)$ and the firm's profit-maximizing condition. □

Proposition 4 underscores the central role of profit-driven firm behavior in shaping the long-term equilibrium level of human employment in AGI-dominated economies. When AGI capital grows without bound and wages collapse toward zero, firms optimally reduce human labor to its irreducible minimum level $\bar{L}_h$, which represents essential tasks that cannot be automated. In this scenario, human work becomes purely residual and economically marginalized, reducing workers to a technologically determined floor of participation. By contrast, when AGI capital growth is bounded and wages stabilize at a positive level, firms retain incentives to employ a larger workforce. The equilibrium level of human labor L_h^∞ in this case reflects a balance between substitution away from labor and the profitability of maintaining human input at a nontrivial scale. Economically, this implies that limits on AGI expansion preserve not only wages but also employment opportunities. From the standpoint of the social contract, this result is particularly salient. Sustained human employment supports social inclusion, political participation, and civic equality, all of which are threatened in scenarios where labor becomes entirely dispensable. Thus, this proposition highlights how macroeconomic and technological constraints on AGI growth may indirectly serve as safeguards of societal cohesion and the normative commitments underlying democratic governance.

3.1 Redistribution, Human Power, and Demand Stabilization: Core Results

The preceding results show that without redistribution, unbounded AGI capital growth leads to the collapse of human wages and economic power. However, technology alone does not determine social outcomes. Institutional choices—especially redistribution through AGI capital taxation and Universal Basic Income—can reshape long-run equilibria. The following propositions formalize this mechanism. First, they show that taxing AGI returns and redistributing income through UBI can prevent human labour driven economic power from collapsing, even when AGI grows without bound. Second, they demonstrate that when AGI growth is bounded, redistribution further strengthens human economic participation, though it becomes less essential. Finally, they establish that without redistribution, declining labor income triggers aggregate demand collapse, threatening economic stability. Together, these results underscore the central role of redistribution in ensuring that technological progress does not marginalize human populations, but instead supports inclusive and stable economic systems.

Proposition 5 (Positive Stabilization of Human Economic Power under AGI Redistribution)**.** *Suppose $\theta > \delta_{AGI}$ and a constant fraction $\tau \in (0,1)$ of AGI-generated capital income is taxed and redistributed equally as Universal Basic Income to humans. Define human disposable income share as*

$$P_h^{UBI}(t) = \frac{w_h(t)L_h(t) + T(t)}{w_h(t)L_h(t) + (1-\tau)r_{AGI}K_{AGI}(t) + r_K K}, \tag{25}$$

where $T(t)$ denotes total UBI transfer income. Then, as $t \to \infty$, $P_h^{UBI}(t)$ converges to a strictly positive limit given by

$$\lim_{t\to\infty} P_h^{UBI}(t) = \frac{\tau}{1-\tau} > 0. \tag{26}$$

Proof. Define human economic power with UBI as the share of total disposable income received by humans

$$P_h^{UBI}(t) = \frac{w_h(t)L_h(t) + T(t)}{w_h(t)L_h(t) + (1-\tau)r_K K + (1-\tau)r_{AGI}K_{AGI}(t)},$$

where $T(t)$ is total UBI transfer, funded by a tax τ on capital income, given by

$$T(t) = \tau(r_K K + r_{AGI}K_{AGI}(t)).$$

By assumption, as $t \to \infty$, $K_{AGI}(t) \to \infty$ (due to $\theta > \delta_{AGI}$), and $w_h(t)L_h(t) \to 0$ by Proposition 2. Thus, in the long run, both total income and UBI transfers are dominated by AGI capital returns. Substituting asymptotic behavior

$$P_h^{UBI}(t) \sim \frac{T(t)}{(1-\tau)r_{AGI}K_{AGI}(t)},$$

since $w_h(t)L_h(t) \to 0$ and K is constant. Using $T(t) = \tau r_{AGI}K_{AGI}(t)$ asymptotically (since $r_K K$ is negligible compared to $r_{AGI}K_{AGI}(t)$), we have

$$P_h^{UBI}(t) \sim \frac{\tau r_{AGI}K_{AGI}(t)}{(1-\tau)r_{AGI}K_{AGI}(t)} = \frac{\tau}{1-\tau}.$$

Since $\tau \in (0,1)$, it follows that

$$\lim_{t\to\infty} P_h^{UBI}(t) = \frac{\tau}{1-\tau} > 0.$$

Thus, human economic power stabilizes at a strictly positive level determined by the tax rate τ. □

Proposition 5 reveals a critical normative implication: in a future economy dominated by AGI capital, where human wages tend toward zero, redistribution mechanisms such as UBI funded by AGI taxation are essential to preserve human labour driven economic relevance. Without redistribution, human labour based economic power would collapse entirely, violating the core principles of the social contract, which is grounded in shared participation and reciprocal recognition among citizens. By stabilizing income and maintaining a positive share of aggregate resources for humans, redistribution ensures that individuals retain not only material security but also the capacity for civic engagement and political agency. In this sense, AGI taxation and redistribution are not merely economic tools but become foundational to renewing the social contract in an era of extreme automation.

Proposition 6 (Effectiveness of UBI under Bounded and Unbounded AGI Growth)**.** *Suppose a constant fraction $\tau \in (0,1)$ of AGI capital income is taxed and redistributed as Universal Basic Income (UBI) to humans. Then*

(i) If AGI capital grows unboundedly ($K_{AGI}(t) \to \infty$), UBI is necessary and sufficient to stabilize human economic power at a strictly positive level, with

$$\lim_{t\to\infty} P_h^{UBI}(t) = \frac{\tau}{1-\tau}. \tag{27}$$

Without UBI, $\tau = 0$, human economic power collapses to zero.

(ii) If AGI capital converges to a finite level $K_{AGI}(t) \to K_{AGI}^{\infty} < \infty$, human economic power remains strictly positive even without UBI, but UBI further increases the steady-state share of income accruing to humans.

Proof. Case (i): Suppose $K_{AGI}(t) \to \infty$. From Proposition 1, $K_{AGI}(t)$ grows without bound, and from Proposition 2 , human labor income $w_h(t)L_h(t) \to 0$ as $t \to \infty$. Thus, without UBI (i.e., if $\tau = 0$), human economic power $P_h(t)$ collapses to zero. However, with a redistribution policy at rate $\tau > 0$, Proposition 5 implies that

$$\lim_{t\to\infty} P_h^{UBI}(t) = \frac{\tau}{1-\tau} > 0,$$

thereby stabilizing human economic power at a strictly positive level determined solely by τ. Case (ii): Suppose $K_{AGI}(t) \to K_{AGI}^{\infty} < \infty$. Then, from Proposition 4, human labor income $w_h(t)L_h(t)$ remains strictly positive in the long run. Therefore, even without UBI (i.e., $\tau = 0$), human economic power $P_h(t)$ converges to a positive finite value.

When UBI is introduced ($\tau > 0$), the redistributed capital income further augments human income, thereby strictly increasing $P_h(t)$ relative to the no-UBI case. Thus, UBI is *necessary* for the preservation of human economic power under unbounded AGI growth, and *beneficial*, though not strictly necessary, under bounded AGI growth. □

Proposition 6 highlights that the necessity of redistribution critically depends on the trajectory of AGI capital growth. When AGI capital expands without bound, human labor becomes economically irrelevant and only redistribution (such as UBI) prevents human economic power from collapsing. In this case, redistribution is indispensable to uphold the social contract, which demands that all members of society retain a meaningful share in economic life. By contrast, if AGI capital growth is bounded, human labor retains a residual productive role, and while redistribution is not strictly necessary for survival, it enhances equity and reinforces social solidarity. Thus, UBI serves a dual role: as an essential lifeline for preserving human agency under extreme automation, and as a mechanism to strengthen the social contract and civic equality even under moderate technological change.

Proposition 7 (Collapse of Human Economic Power Without Redistribution)**.** *Suppose $\tau = 0$ and $K_{AGI}(t) \to \infty$. Then human economic power $P_h(t)$ collapses to zero*

$$\lim_{t\to\infty} P_h(t) = 0. \tag{28}$$

Proof. Recall that human economic power

$$P_h(t) = \frac{w_h(t)L_h(t)}{w_h(t)L_h(t) + r_K K + r_{AGI}K_{AGI}(t)}.$$

From Proposition 3, we know that $w_h(t) \to 0$ as $t \to \infty$, and $L_h(t)$ is bounded (since $L_h(t) \geq \bar{L}_h > 0$ by technological constraint). Therefore

$$\lim_{t\to\infty} w_h(t)L_h(t) = 0.$$

Since $K_{AGI}(t) \to \infty$ by assumption, and $r_{AGI} > 0$ by definition, it follows that

$$\lim_{t\to\infty} r_{AGI}K_{AGI}(t) = +\infty.$$

Meanwhile, $r_K K$ is constant over time. Thus, the total denominator

$$w_h(t)L_h(t) + r_K K + r_{AGI}K_{AGI}(t)$$

diverges to $+\infty$ as $t \to \infty$. Now, the numerator tends to zero and the denominator tends to infinity. Thus, applying basic limit laws

$$\lim_{t\to\infty} P_h(t) = \lim_{t\to\infty} \frac{w_h(t)L_h(t)}{w_h(t)L_h(t) + r_K K + r_{AGI}K_{AGI}(t)} = 0.$$

Therefore, human economic power collapses to zero asymptotically

$$\lim_{t\to\infty} P_h(t) = 0.$$

□

Proposition 7 demonstrates that in the absence of redistribution mechanisms, human economic power inevitably collapses as AGI capital grows without bound. With labor income vanishing and capital income concentrated entirely among AGI owners, humans become economically marginalized. This outcome not only reflects extreme inequality but fundamentally violates the principles of the social contract, which requires that all individuals retain meaningful participation and agency within the economic and political order. Without redistribution, the economy drifts toward a form of digital feudalism, where economic production flourishes but civic equality and shared prosperity disintegrate. Thus, in an AGI-dominated economy, redistribution is not merely a tool of economic policy but a political necessity to preserve the legitimacy and cohesion of the social contract itself.

Proposition 8 (Steady-State Human Power as a Function of Redistribution Rate)**.** *Suppose $K_{AGI}(t) \to \infty$ and $\tau \in (0,1)$. Then the steady-state human economic power under UBI satisfies*

$$\lim_{t\to\infty} P_h^{UBI}(t) = \frac{\tau}{1-\tau}, \tag{29}$$

and the function $\tau \mapsto \frac{\tau}{1-\tau}$ is strictly increasing and continuous on $(0,1)$.

Proof. Recall that under UBI, the human economic power is defined as

$$P_h^{UBI}(t) = \frac{w_h(t)L_h(t) + T(t)}{w_h(t)L_h(t) + r_K K + r_{AGI}K_{AGI}(t)},$$

where $T(t) = \tau r_{AGI}K_{AGI}(t)$ denotes the total tax-funded transfer to humans, with $\tau \in (0,1)$ the redistribution rate. Since $K_{AGI}(t) \to \infty$ as $t \to \infty$, and recalling from Proposition 3 that $w_h(t)L_h(t) \to 0$, the asymptotic behavior of each term is

$$\begin{aligned} w_h(t)L_h(t) &\to 0, \\ r_K K &\text{ remains constant,} \\ r_{AGI}K_{AGI}(t) &\to \infty, \\ T(t) = \tau r_{AGI}K_{AGI}(t) &\to \infty. \end{aligned}$$

Thus, asymptotically, we can neglect the terms $w_h(t)L_h(t)$ and $r_K K$ compared to $r_{AGI}K_{AGI}(t)$, and approximate

$$P_h^{UBI}(t) \sim \frac{T(t)}{r_{AGI}K_{AGI}(t)} \quad \text{as} \quad t \to \infty.$$

Substituting $T(t) = \tau r_{AGI}K_{AGI}(t)$ gives

$$P_h^{UBI}(t) \sim \frac{\tau r_{AGI}K_{AGI}(t)}{r_{AGI}K_{AGI}(t)} = \tau,$$

but recall that the denominator in the original $P_h^{UBI}(t)$ includes only the untaxed portion of capital income $r_{AGI}(1-\tau)K_{AGI}(t)$ in addition to labor income. Thus, the correct asymptotic formula is

$$P_h^{UBI}(t) \sim \frac{\tau r_{AGI}K_{AGI}(t)}{(1-\tau)r_{AGI}K_{AGI}(t)} = \frac{\tau}{1-\tau}.$$

Therefore,

$$\lim_{t\to\infty} P_h^{UBI}(t) = \frac{\tau}{1-\tau}.$$

The function

$$f(\tau) = \frac{\tau}{1-\tau}$$

is defined on the open interval $\tau \in (0,1)$. Differentiating f with respect to τ gives

$$f'(\tau) = \frac{1}{(1-\tau)^2} > 0 \quad \text{for all} \quad \tau \in (0,1).$$

It follows that f is strictly increasing on $(0,1)$. Moreover, $f(\tau)$ is continuous on $(0,1)$, as it is the ratio of two continuous functions with non-vanishing denominator. Hence, the steady-state human economic power under UBI is given by $\frac{\tau}{1-\tau}$, which is strictly increasing and continuous in τ on the domain $(0,1)$. □

Proposition 8 reveals that in an AGI-dominated economy, human labour driven economic power becomes a direct and continuous function of political and institutional choices—specifically, the redistribution rate τ. The result implies that the degree to which society taxes AGI capital and redistributes income fundamentally determines the extent to which humans retain economic agency. A higher τ ensures proportionally greater human economic power, while a lower τ erodes it. This establishes redistribution not merely as a fiscal tool, but as a foundational mechanism for renewing the social contract in the face of automation. Without adequate redistribution, humans risk exclusion from the benefits of economic progress, undermining civic equality and collective legitimacy. In contrast, setting τ sufficiently high can restore a meaningful stake for all individuals, aligning AGI-driven prosperity with the normative commitments of social and political inclusion.

Proposition 9 (Persistence of Human Economic Power under Bounded AGI Capital)**.** *Suppose that $K_{AGI}(t) \to K_{AGI}^{\infty} < \infty$ as $t \to \infty$ and that the redistribution rate satisfies $\tau = 0$. Assume further that $L_h(t) \geq \bar{L}_h > 0$ for all t, and that the human wage $w_h(t)$ converges to a strictly positive steady-state value $w_h^{\infty} > 0$. Then, human labor income $w_h(t)L_h(t)$ remains strictly positive in the limit, and the corresponding measure of human economic power*

$$P_h(t) = \frac{w_h(t)L_h(t)}{w_h(t)L_h(t) + r_K K + r_{AGI} K_{AGI}(t)} \tag{30}$$

converges to a strictly positive finite limit as $t \to \infty$

$$\lim_{t\to\infty} P_h(t) = P_h^{\infty} > 0. \tag{31}$$

Proof. Recall that human economic power without UBI is defined by

$$P_h(t) = \frac{w_h(t)L_h(t)}{w_h(t)L_h(t) + r_K K + r_{AGI} K_{AGI}(t)}.$$

Since $K_{AGI}(t) \to K_{AGI}^{\infty} < \infty$, AGI capital stock converges to a finite positive value. From Proposition 3, we know that

$$\lim_{t\to\infty} w_h(t) = w_h^{\infty} > 0,$$

and $L_h(t) \geq \bar{L}_h > 0$ for all t, implying that

$$\liminf_{t\to\infty} L_h(t) \geq \bar{L}_h > 0.$$

Thus, their product satisfies

$$\liminf_{t\to\infty} w_h(t)L_h(t) \geq w_h^{\infty}\bar{L}_h > 0,$$

which guarantees that human labor income remains strictly positive in the limit. Moreover, the denominator of $P_h(t)$,

$$w_h(t)L_h(t) + r_K K + r_{AGI} K_{AGI}(t),$$

also converges to a finite strictly positive value, because $w_h(t)L_h(t)$ converges to a positive number, $r_K K$ is constant and positive, and $r_{AGI}K_{AGI}(t)$ converges to $r_{AGI}K_{AGI}^{\infty} > 0$. Thus, both the numerator and denominator converge to finite strictly positive values as $t \to \infty$.

Therefore, the ratio $P_h(t)$ converges to a finite strictly positive limit:

$$\lim_{t\to\infty} P_h(t) > 0.$$

□

Proposition 9 demonstrates that when AGI capital accumulation is naturally bounded by technological or economic constraints, human labor retains a durable and nontrivial role in the economy, even in the absence of redistributive mechanisms such as UBI. The finite saturation of AGI capital ensures that wages and employment stabilize at strictly positive levels, preserving a nonzero share of aggregate income and preventing total economic exclusion of human workers. This represents a "soft disruption" scenario, where technological advancement does not fully displace human agency. However, while survival is thus mathematically guaranteed, the share of total income accruing to labor may be substantially diminished, threatening civic equality and social cohesion. From the standpoint of the social contract, natural limits on AGI growth act as a structural safeguard for inclusion, but do not suffice to ensure fairness, dignity, and equal participation. Therefore, although bounded AGI growth avoids the extreme marginalization seen in unbounded cases, deliberate redistributive policies may still be essential to renew the social contract and uphold its deeper normative commitments in an AI-driven economy.

Proposition 10 (Collapse of Human Economic Power Without Redistribution)**.** *Suppose $\tau = 0$ and $K_{AGI}(t) \to \infty$. Then human economic power $P_h(t)$ collapses to zero*

$$\lim_{t\to\infty} P_h(t) = 0. \tag{32}$$

Proof. Recall that human economic power without redistribution is defined by

$$P_h(t) = \frac{w_h(t)L_h(t)}{w_h(t)L_h(t) + r_K K + r_{AGI}K_{AGI}(t)}.$$

Since $\tau = 0$, there are no transfers $T(t)$, and human income consists solely of labor income $w_h(t)L_h(t)$. From Proposition 2, we have

$$\lim_{t\to\infty} w_h(t)L_h(t) = 0.$$

Meanwhile, by assumption, $K_{AGI}(t) \to \infty$, and $r_{AGI} > 0$ by definition, thus

$$\lim_{t\to\infty} r_{AGI}K_{AGI}(t) = +\infty.$$

Also, $r_K K$ is a constant finite term. Thus, the denominator

$$w_h(t)L_h(t) + r_K K + r_{AGI}K_{AGI}(t)$$

diverges to $+\infty$ as $t \to \infty$. Hence, the fraction $P_h(t)$ is of the form "zero over infinity," leading by standard limit properties to

$$\lim_{t\to\infty} P_h(t) = 0.$$

□

Proposition 10 reveals the stark consequence of failing to implement redistributive policies in an economy dominated by unbounded AGI growth. As AGI capital expands and replaces labor, human wages and employment collapse, and with them, human economic power vanishes entirely. This outcome reflects more than extreme inequality—it signifies the complete marginalization of humans from economic life. From the perspective of the social contract, such exclusion constitutes a fundamental breakdown: the promise of mutual obligation, equality, and shared participation in collective prosperity dissolves when a segment of society—humans—no longer holds meaningful economic agency. Without mechanisms like UBI to recycle AGI-generated wealth, technological progress thus becomes incompatible with the social contract's normative ideals, threatening to transform a market economy into a form of digital feudalism where humans exist outside the domain of economic power and political voice.

Proposition 11 (Demand Collapse without UBI)**.** *Suppose $\tau = 0$. If $w_h(t) \to 0$ as $t \to \infty$, then aggregate demand $D(t) \to 0$ as $t \to \infty$.*

Proof. Aggregate demand is given by

$$D(t) = c_h\left(w_h(t)L_h(t)\right) + c_K\left(r_K K\right) + c_{AGI}\left(r_{AGI}K_{AGI}(t)\right),$$

where $c_h \in (0,1)$ is the marginal propensity to consume (MPC) out of labor income, and $c_K, c_{AGI} \in (0,1)$ are the MPCs out of traditional capital income and AGI capital income, respectively. Since $\tau = 0$, there are no transfers (no UBI); all consumption depends only on private incomes. We analyze each term

Step 1: Behavior of $c_h w_h(t) L_h(t)$. From Proposition 2, $w_h(t) \to 0$ as $t \to \infty$, and $L_h(t)$ is bounded below by $\bar{L}_h > 0$. Thus,

$$\lim_{t\to\infty} w_h(t) L_h(t) = 0,$$

implying

$$\lim_{t\to\infty} c_h w_h(t) L_h(t) = 0.$$

Step 2: Behavior of capital incomes. $r_K K$ is constant and finite. $r_{AGI} K_{AGI}(t) \to +\infty$ under unbounded AGI capital growth (Proposition 1). However, we note that: (i) The MPCs c_K and c_{AGI} are typically much lower than c_h, because capital owners have a lower propensity to consume (wealthier agents save more).(ii) In the model, it is assumed that $c_K, c_{AGI} \ll 1$, i.e., capital owners save a large share of their income. Thus, even though $r_{AGI} K_{AGI}(t)$ grows, the induced consumption from AGI capital owners is small relative to output

$$c_{AGI} r_{AGI} K_{AGI}(t) \ll Y(t).$$

Step 3: Asymptotic behavior of $D(t)$. Since (i) The contribution from human labor consumption collapses to zero, and (ii) the contribution from capital consumption remains negligible relative to output because of low c_K and c_{AGI}, it follows that aggregate demand $D(t)$ falls toward zero relative to total output $Y(t)$. In absolute terms, since the overwhelming mass of the economy shifts to AGI capital whose owners consume only a small fraction of income, and humans' disposable income vanishes, aggregate demand $D(t) \to 0$ as $t \to \infty$. □

Proposition 11 highlights a fundamental instability in an AGI-driven economy without redistribution. As human wages vanish and labor income collapses, the main source of broad-based consumption disappears. While AGI capital income expands, it accrues to owners with low marginal propensities to consume, resulting in insufficient aggregate demand. This imbalance leads to an economic paradox: even as productive capacity grows without limit, the economy faces stagnation or contraction due to demand shortfalls. From a social contract perspective, this outcome is deeply troubling. It signals not only growing inequality but also a breakdown of the reciprocal obligations and participation that legitimize economic and political systems. When a large share of the population loses meaningful economic agency and consumption power, civic inclusion erodes, social cohesion weakens, and the foundational promises of shared prosperity and mutual benefit collapse. Thus, redistribution through mechanisms like UBI is not merely a matter of fairness—it becomes essential for preserving macroeconomic stability and sustaining the legitimacy of the social contract itself.

The results presented jointly establish the conditions under which human economic power can persist or collapse in the transition toward an AGI-dominated economy. They highlight the pivotal roles of AGI capital growth dynamics and redistributive institutional mechanisms. The key findings can be summarized as follows: If AGI capital grows without bound ($K_{AGI}(t) \to \infty$) and no redistribution policy is implemented ($\tau = 0$), human labor income $w_h(t)L_h(t)$ collapses to zero, and human economic power $P_h(t)$ converges to zero. Consequently, humans are asymptotically marginalized from the economy (Propositions 10 and 7). If a constant fraction $\tau \in (0, 1)$ of AGI capital income is taxed and redistributed as UBI, human economic power stabilizes at a strictly positive steady-state value given by $\frac{\tau}{1-\tau}$. Thus, UBI is both necessary and sufficient to preserve positive human economic agency under unlimited AGI growth (Propositions 5 and 8). If AGI capital accumulation is bounded ($K_{AGI}(t) \to K_{AGI}^{\infty} < \infty$), human labor income remains strictly positive even without redistribution. Therefore, human economic power $P_h(t)$ converges to a positive steady-state value, although it may be diminished relative to historical norms (Proposition 9). Although UBI is not strictly necessary when AGI capital is bounded, introducing UBI still increases human economic power by augmenting disposable income, thereby improving equity and economic resilience (Proposition 6). In the absence of redistribution, as human labor income collapses, aggregate demand $D(t)$ also collapses to zero due to the low marginal propensity to consume of capital owners. This threatens macroeconomic instability despite the technical capability for unbounded production growth (Proposition 11). Together, these results demonstrate that the sustainability of human economic participation in an AGI economy crucially depends not only on technological factors but also on institutional arrangements—particularly the design of ownership and redistribution policies. Without reforms such as AGI capital taxation and universal redistribution, the long-run equilibrium is one of extreme concentration of wealth and marginalization of the human population, irrespective of technological prosperity.

4 Analysis of the Extended AGI Model

In the baseline model, AGI capital accumulated without explicit regard for operating and scaling costs, implying potentially unbounded self-improvement and accumulation dynamics. However, in reality, expanding complex technologies such as AGI entails increasing operational, maintenance, and coordination burdens. This extended model introduces these critical frictions by incorporating fixed and variable cost functions into the AGI capital accumulation process. Fixed costs capture infrastructure, regulatory, and organizational overhead, while variable costs depend on the current scale of AGI capital and are classified as sublinear, linear, or superlinear. This extension allows the model to analyze whether AGI capital stabilizes naturally or grows without bound.

We consider the dynamic accumulation of AGI capital $K_{AGI}(t)$ under endogenous investment and cost structures. The law of motion for $K_{AGI}(t)$ is governed by the following differential equation

$$\dot{K}_{AGI}(t) = \phi s_R Y(t) + (\theta - \delta_{AGI})K_{AGI}(t) - C(K_{AGI}(t), F(t)), \tag{33}$$

where $Y(t)$ denotes total economic output, $s_R \in (0,1)$ is the share allocated to AGI-directed R&D, $\phi > 0$ measures the productivity of that R&D, $\theta > 0$ is the endogenous self-improvement rate of AGI, and $\delta_{AGI} > 0$ denotes depreciation.

The total cost function $C : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is assumed separable into fixed and variable components

$$C(K, F) = F(t) + V(K), \tag{34}$$

where $F(t)$ represents fixed operating costs and $V(K)$ captures variable costs that depend on the current capital stock. The fixed cost path $F(t)$ is modeled as piecewise constant and right-continuous. Formally, there exists a sequence $\{F_n\}_{n\in\mathbb{N}_0} \subset \mathbb{R}_+$ such that

$$F(t) = F_n, \quad \text{for all } t \in [5n, 5(n+1))^{1}. \tag{35}$$

We impose the following regularity assumptions on the variable cost function $V : \mathbb{R}_+ \to \mathbb{R}_+$. $V(K) > 0$ for all $K > 0$ (strict positivity), $V'(K) > 0$ for all $K > 0$ (strict monotonicity), and $V''(K) \geq 0$ for all $K > 0$ (weak convexity). To classify the long-run behavior of costs, we define

Definition 1 (Superlinear Cost Function)**.** *We say that V is superlinear if there exists $\epsilon > 0$ such that*

$$\liminf_{K\to\infty} \frac{V(K)}{K^{1+\epsilon}} > 0. \tag{36}$$

Definition 2 (Linear Cost Function)**.** *A function V is linear if there exists $\ell > 0$ such that*

$$\lim_{K\to\infty} \frac{V(K)}{K} = \ell. \tag{37}$$

Definition 3 (Sublinear Cost Function)**.** *A function V is sublinear if*

$$\lim_{K\to\infty} \frac{V(K)}{K} = 0. \tag{38}$$

The positivity of $V(K)$ reflects the fundamental resource requirements of AGI operation—energy, computation, supervision—at all scales. Monotonicity expresses the notion that larger systems impose higher coordination, maintenance, and operational burdens. Convexity captures systemic fragility or congestion effects, where increasing scale introduces disproportionate marginal difficulties. Superlinear growth arises when such effects dominate at high K, leading to endogenous limits on scalability. These properties are broadly consistent with empirical patterns in high-tech infrastructure, cloud systems, and complex organizational forms.

A stationary equilibrium satisfies $\dot{K}_{AGI}(t) = 0$. Letting all variables approach their long-run limits, we define the steady-state level K_{AGI}^{∞} as the solution to the nonlinear algebraic condition

$$(\theta - \delta_{AGI})K_{AGI}^{\infty} + \phi s_R Y^{\infty} = F^{\infty} + V(K_{AGI}^{\infty}), \tag{39}$$

where $F^{\infty} := \lim_{n\to\infty} F_n$ and $Y^{\infty} := \lim_{t\to\infty} Y(t)$. The existence and uniqueness of K_{AGI}^{∞} depend on the relative curvature of the LHS and RHS in the equation. In particular, if $V(K)$ is superlinear, then a unique and finite solution typically exists due to the faster growth of costs relative to productivity.

Theorem 1 (Transition Between Bounded and Unbounded AGI Growth with Stepwise Fixed Costs)**.** *Suppose the AGI capital evolves as above, with $V(\cdot)$ satisfying positivity, monotonicity, and optional convexity. Then*

[1] We assume that fix costs adjust every 5 "time periods". This can be adjusted at no costs.

(i) If $V(K_{AGI}) = \Omega(K_{AGI}^{1+\epsilon})$ for some $\epsilon > 0$, then $K_{AGI}(t)$ remains bounded as $t \to \infty$.

(ii) If $V(K_{AGI}) = o(K_{AGI})$ and aggregate demand $D(t)$ does not collapse, then $K_{AGI}(t) \to \infty$.

Moreover, stepwise increases in $F(t)$ impose short-run rigidity but do not prevent the long-run asymptotic classification.

Proof. The AGI capital accumulation dynamic is given by

$$\dot{K}_{AGI}(t) = \phi s_R Y(t) + (\theta - \delta_{AGI})K_{AGI}(t) - F(t) - V(K_{AGI}(t)).$$

Asymptotically, $F(t) \to F^\infty$ and $Y(t) \to Y^\infty$ (bounded by assumption and aggregate demand constraints). Thus, for large t, the dynamic reduces to

$$\dot{K}_{AGI}(t) \approx C_1 + (\theta - \delta_{AGI})K_{AGI}(t) - V(K_{AGI}(t)),$$

where

$$C_1 := \phi s_R Y^\infty - F^\infty$$

is a constant. We now analyze the long-run dynamics based on the asymptotic growth of $V(K_{AGI})$.

Case (i): Superlinear Costs ($V(K) = \Omega(K^{1+\epsilon})$ for some $\epsilon > 0$). By definition, there exist constants $K_0 > 0$ and $M > 0$ such that for all $K > K_0$,

$$V(K) \geq MK^{1+\epsilon}.$$

Meanwhile, the linear term $(\theta - \delta_{AGI})K$ grows proportionally to K. Therefore, for large K,

$$\dot{K}_{AGI}(t) \leq C_1 + (\theta - \delta_{AGI})K - MK^{1+\epsilon}.$$

As $K \to \infty$, the superlinear term $-MK^{1+\epsilon}$ dominates, so

$$\dot{K}_{AGI}(t) \to -\infty.$$

Thus, $K_{AGI}(t)$ cannot diverge and must eventually decrease when sufficiently large. Since $K_{AGI}(t) \geq 0$ by definition, it follows that

$$K_{AGI}(t) \text{ remains bounded as } t \to \infty.$$

Case (ii): Sublinear Costs ($V(K) = o(K)$). By definition, for any $\epsilon > 0$, there exists K_1 such that for all $K > K_1$,

$$V(K) \leq \epsilon K.$$

Substituting into the dynamic equation for $K > K_1$,

$$\dot{K}_{AGI}(t) \geq C_1 + (\theta - \delta_{AGI})K - \epsilon K.$$

Choose ϵ sufficiently small so that

$$\theta - \delta_{AGI} - \epsilon > 0.$$

Assuming $C_1 \geq 0$ (non-collapsing demand), the right-hand side becomes positive and grows linearly in K. Therefore,

$$\dot{K}_{AGI}(t) > 0$$

for large K, implying that

$$K_{AGI}(t) \to \infty \text{ as } t \to \infty.$$

Stepwise Fixed Costs. Stepwise increases in $F(t)$ introduce discrete downward shocks to $\dot{K}_{AGI}(t)$ but do not affect the asymptotic classification:

- They can slow down or temporarily reverse AGI capital growth.
- However, they do not change whether $K_{AGI}(t)$ is ultimately bounded or unbounded, which depends solely on the asymptotic behavior of $V(K)$.

Hence, we have shown that if $V(K)$ is superlinear, AGI capital remains bounded. If $V(K)$ is sublinear and demand does not collapse, AGI capital diverges. Stepwise fixed costs affect transitional dynamics but not the long-run classification. □

Theorem 1 highlights a fundamental distinction between technological dynamics with and without endogenous friction. When AGI variable costs grow superlinearly, the economy naturally limits AGI accumulation: as expansion becomes increasingly costly, growth stalls, and AGI capital stabilizes at a finite level. This guarantees that human labor retains a non-negligible role in production, since AGI cannot fully displace human input at scale. In contrast, if AGI costs are sublinear, unchecked capital accumulation occurs. AGI grows without bound, eventually outcompeting human labor, driving wages toward zero, and eroding human economic power. From the perspective of the social contract, the presence of superlinear costs acts as a structural safeguard for human economic agency. Natural limits on AGI scalability prevent the technological marginalization of humans, preserving a baseline of inclusion in economic production and income distribution. Without such frictions, however, market forces alone drive toward extreme inequality and social exclusion. Thus, the shape of AGI cost structures is not merely a technological feature — it is a determinant of societal stability. In this sense, regulating or engineering such frictions becomes essential to upholding the normative ideals of the social contract: fairness, dignity, and participation in the shared prosperity of the economy.

Theorem 2 (Existence and Uniqueness of Steady-State AGI Capital)**.** *Suppose the AGI capital accumulation follows*

$$\dot{K}_{AGI}(t) = \phi s_R Y(t) + (\theta - \delta_{AGI})K_{AGI}(t) - F(t) - V(K_{AGI}(t)), \tag{40}$$

where $V(K_{AGI})$ satisfies $V(K_{AGI}) > 0$, $V'(K_{AGI}) > 0$, and $V''(K_{AGI}) \geq 0$ for all $K_{AGI} > 0$, $F(t) \to F^\infty > 0$ as $t \to \infty$, and $Y(t) \to Y^\infty > 0$ as $t \to \infty$. Then there exists a unique finite steady-state $K^\infty_{AGI} > 0$ solving

$$(\theta - \delta_{AGI})K^\infty_{AGI} + \phi s_R Y^\infty = F^\infty + V(K^\infty_{AGI}). \tag{41}$$

Proof. Define the steady-state function

$$g(K) := (\theta - \delta_{AGI})K + \phi s_R Y^\infty - F^\infty - V(K).$$

The steady-state values of K correspond to solutions of $g(K) = 0$.

Step 1: Continuity and limits at endpoints. The function $g(K)$ is continuous because it is a sum of continuous functions: linear in K and continuous in $V(K)$. *At* $K = 0$,

$$g(0) = \phi s_R Y^\infty - F^\infty - V(0).$$

This is finite. It may be positive, negative, or zero depending on parameters, but its finiteness is guaranteed. *As* $K \to \infty$, observe that

$$(\theta - \delta_{AGI})K \quad \text{grows linearly, while} \quad V(K) \quad \text{grows at least linearly and possibly faster (convexity).}$$

Since $V(K)$ is strictly increasing and convex, and $V'(K) > 0$, eventually $V(K)$ dominates $(\theta - \delta_{AGI})K$, and

$$\lim_{K\to\infty} g(K) = -\infty.$$

Step 2: Existence. By continuity of $g(K)$ and the Intermediate Value Theorem, and since

$$g(0) = \text{finite (possibly positive)}, \quad \lim_{K\to\infty} g(K) = -\infty,$$

there exists at least one value $K^\infty_{AGI} > 0$ such that

$$g(K^\infty_{AGI}) = 0.$$

Step 3: Uniqueness. Differentiate $g(K)$

$$g'(K) = (\theta - \delta_{AGI}) - V'(K).$$

Since $V'(K) > 0$ everywhere and $V''(K) \geq 0$, $V'(K)$ is weakly increasing and eventually dominates the constant $(\theta - \delta_{AGI})$.

- If $g'(K) < 0$ for all sufficiently large K, then $g(K)$ is strictly decreasing beyond some point.
- Since $g(K)$ crosses zero exactly once when moving from finite (possibly positive) $g(0)$ to $-\infty$, the solution is unique.

Therefore, there exists exactly one finite $K_{AGI}^{\infty} > 0$ solving $g(K) = 0$. □

Theorem 2 shows that when AGI scaling costs grow sufficiently with size—through rising coordination, energy, and maintenance burdens—unbounded accumulation of AGI capital becomes impossible. Instead, a finite steady-state emerges, balancing self-improvement, R&D productivity, and rising operational costs. Economically, this means that AGI cannot fully substitute away human labor, because at large scales AGI becomes increasingly costly relative to its productivity. As a result, the marginal productivity of human labor remains strictly positive in the long run, preserving human wages, employment, and participation in production. Viewed through the lens of the social contract, this natural ceiling on AGI expansion functions as an endogenous safeguard for human agency and inclusion. Without it, unchecked AGI growth would displace human labor entirely, leading to the collapse of labor income and eroding the foundations of democratic economic citizenship. In contrast, bounded AGI accumulation ensures that technological progress remains socially embedded: humans retain not only relevance, but also bargaining power and a share in prosperity. Thus, superlinear cost structures implicitly uphold key pillars of the social contract — fairness, participation, and protection against exclusion — without requiring constant intervention.

Proposition 12 (Positive Stabilization of Human Wages and Employment under Bounded AGI Growth)**.** *Suppose $K_{AGI}(t) \to K_{AGI}^{\infty}$ as guaranteed by Theorem 2, and that output $Y(t)$ aggregates AGI capital and human labor $L_h(t)$ through a CES production function. Then*

(i) Human wages $w_h(t)$ converge to a strictly positive steady-state value $w_h^{\infty} > 0$.

(ii) Human employment $L_h(t)$ converges to a strictly positive steady-state level $L_h^{\infty} > 0$.

(iii) Aggregate demand $D(t)$ stabilizes at a strictly positive level as $t \to \infty$.

Proof. Recall that the human wage is given by the marginal product of labor

$$w_h(t) = \frac{\partial Y(t)}{\partial L_h(t)},$$

where $Y(t)$ depends on $K_{AGI}(t)$ and $L_h(t)$ through a CES aggregator

$$Y(t) = A\left(\delta_K(K + K_{AGI}(t))^{\rho} + \delta_L L_h(t)^{\rho}\right)^{1/\rho}.$$

By Theorem 2, $K_{AGI}(t) \to K_{AGI}^{\infty} < \infty$ as $t \to \infty$. Thus, $K_{AGI}(t)$ becomes asymptotically constant at K_{AGI}^{∞}, and therefore

$$\delta_K(K + K_{AGI}(t))^{\rho} \to \delta_K(K + K_{AGI}^{\infty})^{\rho},$$

a finite, positive constant. Now, since $K_{AGI}(t)$ stabilizes, the CES aggregator effectively behaves like a function of $L_h(t)$ alone in the long run. Specifically, the marginal product of labor becomes

$$w_h(t) = A\left(\delta_K(K + K_{AGI}^{\infty})^{\rho} + \delta_L L_h(t)^{\rho}\right)^{\frac{1}{\rho}-1} \delta_L L_h(t)^{\rho-1}.$$

Observe

- The first term $\left(\delta_K(K + K_{AGI}^{\infty})^{\rho} + \delta_L L_h(t)^{\rho}\right)^{\frac{1}{\rho}-1}$ is strictly positive and continuous for all $L_h(t) > 0$.
- The second term $\delta_L L_h(t)^{\rho-1}$ is bounded away from zero if $L_h(t)$ remains bounded away from zero.

Step 1: Behavior of $L_h(t)$. Firms maximize profits by hiring $L_h(t)$ to satisfy the first-order condition

$$\frac{\partial Y(t)}{\partial L_h(t)} = w_h(t).$$

Given that $w_h(t)$ is determined competitively, and labor cannot be arbitrarily reduced below a minimum technological bound $\bar{L}_h > 0$ (by earlier model assumptions), it follows that $L_h(t)$ must converge to some steady-state level $L_h^{\infty} \geq \bar{L}_h > 0$. Moreover, since $w_h(t)$ converges to match the strictly positive marginal product at L_h^{∞}, $w_h(t)$ converges to a strictly positive w_h^{∞}. Thus

$$\lim_{t\to\infty} L_h(t) = L_h^{\infty} > 0, \quad \lim_{t\to\infty} w_h(t) = w_h^{\infty} > 0.$$

Step 2: Behavior of aggregate demand $D(t)$. Aggregate demand is modeled as a function of human disposable income and capital owner income. In particular, since human labor income $w_h(t)L_h(t) \to w_h^\infty L_h^\infty > 0$, and assuming a positive marginal propensity to consume $c_h \in (0,1)$ out of labor income, the contribution of human workers to aggregate demand remains strictly positive. Hence

$$\liminf_{t\to\infty} D(t) > 0,$$

and demand does not collapse. $\square$

Proposition 12 demonstrates that when AGI capital accumulation is naturally bounded, human labor retains a structurally stable and meaningful role in the economy. In particular, finite AGI capital ensures that the marginal productivity of labor remains strictly positive, which in turn sustains human wages and employment at non-trivial steady-state levels. Aggregate demand also stabilizes because workers continue to earn and spend income, preventing macroeconomic collapse. Viewed through the lens of the social contract, this bounded AGI scenario preserves essential elements of economic citizenship. Humans continue to participate as productive contributors rather than becoming economically obsolete. Their wages, employment, and consumption not only provide personal income but also uphold broader social and economic cohesion. This result highlights a key distinction between "soft" and "hard" technological disruption. When AGI growth is limited by cost structures or physical constraints, technology complements rather than fully substitutes for human labor. This preserves the conditions necessary for fairness, reciprocity, and inclusion—core principles of the social contract. By contrast, in scenarios of unbounded AGI expansion, these protections erode, and human exclusion becomes a structural inevitability. Thus, bounded AGI growth acts as a natural institutional ally of the social contract. It prevents full technological displacement, preserves wages and work, and supports the ongoing distribution of income through market mechanisms, reducing reliance on redistributive transfers or political interventions to maintain social stability.

Theorem 3 (Threshold for Viable AGI Operation)**.** *Suppose fixed costs converge to $F^\infty > 0$, and the variable cost $V(K_{AGI})$ satisfies positivity, monotonicity, and convexity. Let AGI capital accumulation satisfy*

$$\dot{K}_{AGI}(t) = (\theta - \delta_{AGI})K_{AGI}(t) + \phi s_R Y(K_{AGI}(t)) - F^\infty - V(K_{AGI}(t)),$$

where $Y(K_{AGI})$ is continuous and satisfies $Y(0) = 0$ with $Y'(K_{AGI}) > 0$ for $K_{AGI} > 0$. Then there exists a minimum viable AGI capital stock $K_{AGI}^{\min} > 0$ such that

$$(\theta - \delta_{AGI})K_{AGI} + \phi s_R Y(K_{AGI}) < F^\infty + V(K_{AGI}) \quad \text{for all} \quad K_{AGI} < K_{AGI}^{\min}. \tag{42}$$

Thus, no sustainable AGI accumulation is possible at subcritical scales $K_{AGI} < K_{AGI}^{\min}$.

Proof. Define the net accumulation function

$$\Gamma(K_{AGI}) := (\theta - \delta_{AGI})K_{AGI} + \phi s_R Y(K_{AGI}) - F^\infty - V(K_{AGI}).$$

We analyze $\Gamma(K_{AGI})$ near $K_{AGI} = 0$. First, observe

$$\Gamma(0) = -F^\infty - V(0) < 0,$$

because $F^\infty > 0$ and $V(0) > 0$ by positivity of variable costs. Next, compute the derivative at $K_{AGI} = 0$:

$$\Gamma'(K_{AGI}) = (\theta - \delta_{AGI}) + \phi s_R Y'(K_{AGI}) - V'(K_{AGI}).$$

At $K_{AGI} = 0$, $Y'(0) > 0$ (since AGI productivity is assumed to rise with scale), and $V'(0) > 0$ by positivity and monotonicity. Thus, $\Gamma'(0)$ may be positive or negative, but what matters is the behavior of Γ in a neighborhood around zero. Because $\Gamma(0) < 0$ and Γ is continuous (as a sum of continuous functions), by continuity there exists $\delta > 0$ such that

$$\Gamma(K_{AGI}) < 0 \quad \text{for all} \quad K_{AGI} \in [0, \delta).$$

Now, since $V(K_{AGI})$ is convex, $F^\infty + V(K_{AGI})$ is convex and increasing, and $(\theta - \delta_{AGI})K_{AGI} + \phi s_R Y(K_{AGI})$ is increasing in K_{AGI} (since both terms are increasing in K_{AGI}). Thus, $\Gamma(K_{AGI})$ eventually becomes non-negative at some $K_{AGI} = K_{AGI}^{\min}$ where

$$\Gamma(K_{AGI}^{\min}) = 0.$$

Moreover, since $\Gamma(K_{AGI})$ transitions from negative to zero, $K_{AGI}^{\min}$ is uniquely defined. Hence, for $K_{AGI} < K_{AGI}^{\min}$, $\Gamma(K_{AGI}) < 0$, and net AGI capital accumulation $\dot{K}_{AGI}(t) < 0$, implying that subcritical scales of AGI cannot sustain themselves. $\square$

Theorem 3 establishes that when AGI systems face nontrivial fixed and variable costs, there exists a minimum viable scale $K_{AGI}^{\min}$ below which AGI capital cannot sustain autonomous growth. At subcritical levels, operational costs exceed productive returns, causing net decumulation and eventual decline of AGI capital unless externally subsidized. Economically, this result implies that small-scale AGI cannot fully displace human labor. When AGI operates below viability thresholds, human workers remain indispensable to production, and wages stay strictly positive. Thus, technological substitution remains incomplete unless and until AGI capital crosses this critical boundary. In relation to the social contract, this structural threshold serves as a protective barrier safeguarding human relevance in the economy. As long as AGI remains below $K_{AGI}^{\min}$, humans continue to play an essential economic role, earning wages, participating in production, and maintaining consumption-driven demand. This preserves the basic conditions for economic inclusion, reciprocity, and social stability — key pillars of the social contract. By contrast, surpassing $K_{AGI}^{\min}$ introduces the possibility of AGI self-sufficiency and large-scale labor displacement. Thus, the viability threshold demarcates a fundamental boundary between two political-economic regimes: one in which humans remain essential productive agents and another in which their economic agency may face erosion. Understanding and regulating this transition zone will be central to any future social contract in AGI-dominated economies.

Proposition 13 (Persistence of Human Labor under AGI Viability Threshold)**.** *Suppose the AGI capital stock $K_{AGI}(t)$ remains strictly below the viability threshold $K_{AGI}^{\min}$ for all t due to initial conditions or persistent shocks. Then*

(i) Human labor $L_h(t)$ remains essential for production.

(ii) Human wages $w_h(t)$ remain strictly positive and bounded away from zero.

Proof. Consider the CES production function

$$Y(t) = A\left(\delta_K (K + K_{AGI}(t))^\rho + \delta_L L_h(t)^\rho\right)^{1/\rho},$$

where $0 < \rho < 1$. Since $K_{AGI}(t) < K_{AGI}^{\min}$ for all t by assumption, and $K_{AGI}^{\min}$ is finite, the capital term $\delta_K(K+K_{AGI}(t))^\rho$ is bounded above by a finite constant

$$\delta_K (K + K_{AGI}(t))^\rho \leq \delta_K (K + K_{AGI}^{\min})^\rho =: \bar{K}.$$

Step 1 (Essentiality of human labor). For total output $Y(t)$ to remain positive and for the economy to operate, it must be that the labor term $\delta_L L_h(t)^\rho$ contributes positively. If $L_h(t) \to 0$, then $\delta_L L_h(t)^\rho \to 0$, and the CES aggregator would asymptotically reduce to

$$Y(t) \to A\bar{K}^{1/\rho},$$

which is finite. However, in equilibrium, firms seek to meet positive output demand $D(t) > 0$, which (as shown in prior results) requires sufficient $Y(t)$. Since capital alone at sub-viability $K_{AGI}(t)$ cannot support indefinitely growing $Y(t)$, labor must be used. Therefore, profit-maximizing firms hire $L_h(t)$ to avoid a reduction in output, implying

$$L_h(t) \geq \bar{L}_h > 0,$$

where $\bar{L}_h$ reflects either technological or economic lower bounds on labor usage.

Step 2 (Boundedness of human wages). The wage rate $w_h(t)$ equals the marginal product of labor

$$w_h(t) = A\left(\delta_K (K + K_{AGI}(t))^\rho + \delta_L L_h(t)^\rho\right)^{\frac{1}{\rho}-1} \cdot \delta_L \rho L_h(t)^{\rho-1}.$$

The first factor remains bounded and strictly positive because:

$$\delta_K (K + K_{AGI}(t))^\rho + \delta_L L_h(t)^\rho \geq \delta_L \bar{L}_h^\rho > 0.$$

The second factor, $\delta_L \rho L_h(t)^{\rho-1}$, is also strictly positive and bounded because $L_h(t) \geq \bar{L}_h > 0$ and $\rho - 1 < 0$. Thus

$$w_h(t) \geq \underline{w}_h > 0,$$

for some finite $\underline{w}_h$, proving that wages remain strictly positive and bounded away from zero. □

Proposition 13 demonstrates that if AGI capital remains below its viability threshold, human labor retains an indispensable role in production. This guarantees the persistence of positive wages and employment, preventing total labor market collapse. From a macroeconomic perspective, this reflects an economy structurally reliant on human inputs, thus preserving broad-based income generation and aggregate demand. In terms of the social contract, this natural technological limit offers a form of "baseline protection." Even absent redistributive policies, humans maintain economic relevance and agency, securing participation in production and access to income. However, while survival and inclusion are assured, fairness and economic equality are not automatic: wage levels and bargaining power may still be eroded relative to pre-AGI eras. Thus, while viability thresholds prevent catastrophic exclusion, complementary institutional arrangements (such as labor standards, inclusive ownership models, or redistributive mechanisms) remain essential to uphold the deeper normative ideals of justice and equal citizenship in an AGI

Theorem 4 (Global Stability with Superlinear Costs)**.** *Suppose the AGI capital accumulation satisfies*

$$\dot{K}_{AGI}(t) = \phi s_R Y(t) + (\theta - \delta_{AGI})K_{AGI}(t) - F^{\infty} - V(K_{AGI}(t)),$$

where

- $V(K_{AGI})$ *is continuous, strictly increasing, convex, and satisfies the superlinearity condition: there exists* $\epsilon > 0$ *such that*
$$\liminf_{K_{AGI}\to\infty} \frac{V(K_{AGI})}{K_{AGI}^{1+\epsilon}} > 0,$$
- $F^{\infty} > 0$ *is the asymptotic fixed cost,*
- $Y(t)$ *is bounded above by some finite* $Y_{\max}$.

Then, regardless of initial condition $K_{AGI}(0) > 0$*, the trajectory* $K_{AGI}(t)$ *remains bounded for all* t *and converges to a unique finite steady-state* $K_{AGI}^{\infty} > 0$ *as* $t \to \infty$.

Proof. We proceed in several steps:

Step 1: Global Boundedness. For large K_{AGI}, by superlinearity of V there exists $M > 0$ and $K_0 > 0$ such that for all $K_{AGI} \geq K_0$

$$V(K_{AGI}) \geq M K_{AGI}^{1+\epsilon}.$$

Meanwhile, $(\theta - \delta_{AGI})K_{AGI}$ grows only linearly in K_{AGI}, and $\phi s_R Y(t) \leq \phi s_R Y_{\max}$ is bounded. Thus, for sufficiently large K_{AGI}, the dominant term in the differential equation is $-V(K_{AGI})$, and we have

$$\dot{K}_{AGI}(t) = \phi s_R Y(t) + (\theta - \delta_{AGI})K_{AGI}(t) - F^{\infty} - V(K_{AGI}(t)) < 0.$$

Therefore, $K_{AGI}(t)$ cannot grow beyond some finite upper bound.

Step 2: Invariance and Compactness. Since $\dot{K}_{AGI}(t)$ is continuous in $K_{AGI}(t)$ and $K_{AGI}(t)$ is bounded above and below ($K_{AGI}(t) \geq 0$), the dynamics are constrained within a compact interval $[0, K_{max}]$ for some finite K_{max}.

Step 3: Existence of Steady-State. Set $\dot{K}_{AGI}(t) = 0$ to find steady states

$$(\theta - \delta_{AGI})K_{AGI}^{\infty} + \phi s_R Y^{\infty} = F^{\infty} + V(K_{AGI}^{\infty}),$$

where Y^{∞} depends on K_{AGI}^{∞} but is bounded. Since V is continuous and strictly increasing, and the left-hand side is continuous and increasing (linear in K_{AGI} plus bounded term), the Intermediate Value Theorem implies a solution exists.

Step 4: Uniqueness of Steady-State. Suppose two distinct steady-states $K_1 < K_2$ exist. Then

- Left-hand side, $(\theta - \delta_{AGI})K$ increases linearly in K.
- Right-hand side, $F^{\infty} + V(K)$ increases faster (by convexity and strict monotonicity of V).

Thus, at $K_2 > K_1$, the right-hand side is larger by a greater amount than the left-hand side (because V grows superlinearly), contradicting steady-state equality at both K_1 and K_2. Thus, the steady-state is unique.

Step 5: Convergence. Since the system is monotone for large K_{AGI} and bounded, and the dynamics push toward the unique steady-state, standard dynamical systems arguments imply that $K_{AGI}(t) \to K_{AGI}^{\infty}$ as $t \to \infty$. □

Theorem 4 establishes that when AGI operating costs grow sufficiently rapidly with scale—specifically, superlinearly—unbounded accumulation becomes infeasible and AGI capital converges to a finite steady-state level. This has profound economic and social implications. First, superlinear costs introduce an inherent economic brake on runaway automation. As AGI capital expands, rising marginal costs eventually outpace linear productivity gains and self-improvement, halting further accumulation. This natural saturation prevents AGI systems from indefinitely expanding their productive capacity and displacing all other factors of production, notably human labor. Second, bounded AGI capital implies that human labor retains an ongoing role in production and income generation. Since AGI capital stabilizes, its marginal contribution to output does not grow without bound. Consequently, the marginal product of human labor—and thus wages—also stabilizes at a strictly positive level, preserving human access to economic resources and participation in market society. Finally, in relation to the social contract, this endogenous stability condition has normative significance. It ensures that, even absent deliberate redistributive interventions, humans are not rendered entirely redundant or economically powerless. Natural technological constraints on AGI expansion thus partially safeguard the principles of inclusion, opportunity, and civic equality that underpin modern conceptions of the social contract. Nonetheless, while exclusion is prevented, fairness is not automatically achieved: labor income shares may still decline significantly. Thus, superlinear costs provide structural protection against extreme forms of economic displacement, but do not obviate the need for complementary institutional measures to realize distributive justice and social cohesion in AGI-intensive economies.

Proposition 14 (Stabilization of Human Wages under Bounded AGI Capital)**.** *Suppose the assumptions of Theorem 4 hold, so that $K_{AGI}(t) \to K_{AGI}^{\infty} < \infty$ as $t \to \infty$. Assume further*

- *The production function $Y(t)$ is CES in $K + K_{AGI}(t)$ and $L_h(t)$,*
- *Human labor supply satisfies $L_h(t) \geq \bar{L}_h > 0$,*
- *The wage $w_h(t)$ is given by the marginal product of human labor*

$$w_h(t) = \frac{\partial Y(t)}{\partial L_h(t)}. \tag{43}$$

Then

$$\lim_{t\to\infty} w_h(t) = w_h^{\infty} > 0, \tag{44}$$

i.e., human wages converge to a strictly positive steady-state value.

Proof. By Theorem 4, $K_{AGI}(t) \to K_{AGI}^{\infty} < \infty$ as $t \to \infty$. Thus, the effective capital stock $K + K_{AGI}(t)$ converges to a finite positive value. The CES production function has the form

$$Y(t) = A\left(\delta_K (K + K_{AGI}(t))^{\rho} + \delta_L L_h(t)^{\rho}\right)^{1/\rho},$$

where $A > 0$, $\delta_K, \delta_L > 0$, and $\rho \in (0, 1)$ (capital and labor are imperfect substitutes). The marginal product of labor (the wage) is

$$w_h(t) = A\left(\delta_K (K + K_{AGI}(t))^{\rho} + \delta_L L_h(t)^{\rho}\right)^{\frac{1}{\rho}-1} \times \delta_L \rho L_h(t)^{\rho-1}.$$

Since $K_{AGI}(t) \to K_{AGI}^{\infty}$ finite, and $L_h(t) \geq \bar{L}_h > 0$ bounded away from zero, the aggregator

$$\delta_K (K + K_{AGI}(t))^{\rho} + \delta_L L_h(t)^{\rho}$$

converges to a finite strictly positive constant. Thus the first term

$$\left(\delta_K (K + K_{AGI}(t))^{\rho} + \delta_L L_h(t)^{\rho}\right)^{\frac{1}{\rho}-1}$$

converges to a finite strictly positive value. The second term

$$\delta_L \rho L_h(t)^{\rho-1}$$

converges to a strictly positive value (because $\bar{L}_h > 0$ and $\rho - 1 < 0$ implies bounded but positive expression). Hence, their product — the wage $w_h(t)$ — converges to a strictly positive constant $w_h^{\infty} > 0$. □

Proposition 14 highlights a fundamental implication of bounded AGI capital for labor markets. When AGI accumulation stabilizes at a finite level due to superlinear cost structures, the marginal product of human labor—and thus wages—also stabilizes at a strictly positive level. Unlike in scenarios of unbounded AGI growth, where capital accumulation indefinitely substitutes for labor and depresses wages toward zero, bounded AGI capital implies that human labor remains an essential and productive input in the economy. Economically, this ensures that humans continue to earn positive wages and retain meaningful participation in market production and consumption. Human labor does not become technologically obsolete, and the risk of complete exclusion from income generation is structurally averted. This is especially important in the context of CES production functions, where imperfect substitutability means that capital cannot entirely replace labor when capital expansion is limited. From the perspective of the social contract, this result is deeply significant. It shows that the inherent frictions in scaling AGI systems—captured here by superlinear cost growth—can serve as natural safeguards of human economic relevance and dignity. Even absent explicit redistributive interventions, humans retain positive wages, enabling them to sustain livelihoods, exercise agency in market interactions, and remain integrated into the economic and civic fabric of society. However, while exclusion is avoided, the fairness of outcomes is not guaranteed: the share of income accruing to labor may still fall relative to capital owners. Thus, bounded AGI growth secures the minimal condition of inclusion essential to the social contract, but distributive justice may still require complementary policies to ensure equitable outcomes.

Theorem 5 (Instability Under Sublinear Cost Structures)**.** *Suppose the variable cost function $V(K_{AGI})$ is sublinear, that is*

$$\lim_{K_{AGI}\to\infty} \frac{V(K_{AGI})}{K_{AGI}} = 0,$$

and that aggregate demand $D(t)$ remains sufficiently non-collapsing, meaning that output $Y(t)$ grows proportionally with $K_{AGI}(t)$ for large t. Then $K_{AGI}(t) \to \infty$ as $t \to \infty$, and no finite steady-state K_{AGI}^{∞} exists.

Proof. Consider the AGI capital accumulation equation

$$\dot{K}_{AGI}(t) = \phi s_R Y(t) + (\theta - \delta_{AGI})K_{AGI}(t) - F(t) - V(K_{AGI}(t)).$$

By assumption, $Y(t)$ grows with $K_{AGI}(t)$, i.e., there exists $\eta > 0$ such that $Y(t) \geq \eta K_{AGI}(t)$ asymptotically, $F(t) \to F^{\infty} > 0$ as $t \to \infty$, and $V(K_{AGI})$ satisfies $\lim_{K_{AGI}\to\infty} \frac{V(K_{AGI})}{K_{AGI}} = 0$. Thus, for any $\epsilon > 0$, there exists $K^* > 0$ such that for all $K_{AGI} > K^*$,

$$V(K_{AGI}) \leq \epsilon K_{AGI}.$$

Substituting into $\dot{K}_{AGI}(t)$, we obtain

$$\dot{K}_{AGI}(t) \geq (\theta - \delta_{AGI} + \phi s_R \eta - \epsilon)K_{AGI}(t) - F^{\infty}.$$

Choose $\epsilon > 0$ sufficiently small so that $\theta - \delta_{AGI} + \phi s_R \eta - \epsilon > 0$. For K_{AGI} large enough such that $(\theta - \delta_{AGI} + \phi s_R \eta - \epsilon)K_{AGI} > 2F^{\infty}$, we then have

$$\dot{K}_{AGI}(t) > \frac{1}{2}(\theta - \delta_{AGI} + \phi s_R \eta)K_{AGI}(t) > 0.$$

Thus, $K_{AGI}(t)$ grows without bound. Therefore,

$$\lim_{t\to\infty} K_{AGI}(t) = +\infty,$$

and no finite steady-state K_{AGI}^{∞} exists. □

Theorem 5 reveals that when AGI operating costs are sublinear in scale, the economy becomes inherently unstable in the long run. As AGI capital expands, self-reinforcing growth dynamics—driven by endogenous improvement and AGI-led output expansion—eventually overwhelm the weakly increasing cost burden. This produces unbounded capital accumulation with no natural stopping point. The social consequences of this regime are stark. As AGI capital grows indefinitely, its contribution to production increasingly displaces human labor, driving down the marginal product of labor and, ultimately, wages. Over time, this process leads to the asymptotic collapse of human earnings and employment opportunities. Without redistributive policies or structural frictions to limit AGI accumulation, human agents risk being pushed out of the economy's core production and distribution systems altogether. From the perspective of the social contract, this scenario represents a severe breakdown. The ideals of inclusion, reciprocity, and shared economic participation—central to any legitimate social order—are undermined when technological forces render human labor redundant and income-less. Sublinear AGI cost structures, therefore, do not merely pose a technical problem of unbalanced growth; they create a fundamental threat to social cohesion and political legitimacy. Addressing this risk may require deliberate institutional interventions to enforce artificial limits or redistribute AGI-generated surpluses to preserve human dignity and economic agency.

Proposition 15 (Collapse of Human Wages Under Sublinear Costs)**.** *Under the assumptions of Theorem 5, if $\rho > 0$ (capital and labor are CES substitutes), then human wages $w_h(t)$ satisfy*

$$\lim_{t\to\infty} w_h(t) = 0, \tag{45}$$

and employment $L_h(t)$ converges to its technological lower bound $\bar{L}_h$.

Proof. From Theorem 5, $K_{AGI}(t) \to \infty$ as $t \to \infty$. Since the production function is CES and capital and labor are substitutes with $\rho > 0$, the marginal product of labor is

$$w_h(t) = A\left(\delta_K(K + K_{AGI}(t))^\rho + \delta_L L_h(t)^\rho\right)^{\frac{1}{\rho}-1} \delta_L L_h(t)^{\rho-1}.$$

As $K_{AGI}(t) \to \infty$, the term $\delta_K(K + K_{AGI}(t))^\rho$ dominates the aggregator, leading the effective marginal product of labor to decay to zero

$$\lim_{t\to\infty} w_h(t) = 0.$$

Given firms minimize labor costs, $L_h(t)$ converges to the minimum feasible level $\bar{L}_h$ imposed by technological constraints. □

Theorem 5 and Proposition 15 jointly reveal the destabilizing consequences of sublinear cost structures for both economic stability and the social contract. When the operating costs of AGI rise too slowly relative to its scale, unchecked accumulation of AGI capital becomes possible. In this regime, AGI-driven production increasingly dominates, while human labor, though technologically indispensable at a minimal level, becomes economically irrelevant. The CES production structure ensures that the marginal product of labor—and thus wages—collapses toward zero, with employment contracting to its bare technological minimum. This outcome poses serious macroeconomic and normative challenges. From an economic perspective, a collapse in labor income erodes aggregate demand and undermines the foundations of inclusive growth, raising the risk of secular stagnation or extreme inequality. From the perspective of the social contract, this trajectory fundamentally violates principles of reciprocity, inclusion, and fair distribution. Humans, once integral contributors to and beneficiaries of economic production, risk being reduced to passive dependents as AGI capital holders capture virtually all economic surplus. Avoiding this dystopian outcome requires either endogenous technological frictions—such as superlinear cost scaling that limits AGI expansion—or explicit institutional interventions, including redistributive taxation, universal transfers, collective ownership, or labor-market protections. These mechanisms can re-anchor humans within the productive core of the economy and preserve economic relevance, civic dignity, and social stability in the face of powerful automation dynamics.

Theorem 6 (Critical Threshold for Human Economic Survival)**.** *Consider the AGI capital accumulation model*

$$\dot{K}_{AGI}(t) = \phi s_R Y(t) + (\theta - \delta_{AGI})K_{AGI}(t) - F(t) - V(K_{AGI}(t)), \tag{46}$$

where $V(\cdot)$ is the variable cost function, $F(t) \to F^\infty > 0$, $Y(t)$ depends positively on $K_{AGI}(t)$, and human labor and AGI capital are CES substitutes with $\rho > 0$. Then

(i) If $V(K_{AGI})$ is superlinear *at infinity (i.e., $\liminf_{K_{AGI}\to\infty} \frac{V(K_{AGI})}{K_{AGI}^{1+\epsilon}} > 0$ for some $\epsilon > 0$), AGI capital remains bounded, human wages stabilize at a positive level, and human economic power persists asymptotically.*

(ii) If $V(K_{AGI})$ is sublinear *(i.e., $\lim_{K_{AGI}\to\infty} \frac{V(K_{AGI})}{K_{AGI}} = 0$) and aggregate demand does not collapse, then $K_{AGI}(t) \to \infty$, human wages collapse to zero, and human economic power vanishes asymptotically.*

Thus, superlinear cost scaling is a necessary and sufficient structural condition for long-run human economic survival without external interventions.

Proof. We proceed in two parts, corresponding to the two cost regimes.

(i) Superlinear Cost Case. Assume $V(K_{AGI})$ satisfies the superlinearity condition, i.e.,

$$\liminf_{K_{AGI}\to\infty} \frac{V(K_{AGI})}{K_{AGI}^{1+\epsilon}} > 0$$

for some $\epsilon > 0$. By Theorem 4, the dynamics of $K_{AGI}(t)$ are globally stable in this regime. Specifically, for sufficiently large K_{AGI}, the negative contribution $-V(K_{AGI})$ dominates the accumulation equation

$$\dot{K}_{AGI}(t) = \phi s_R Y(t) + (\theta - \delta_{AGI})K_{AGI}(t) - F^{\infty} - V(K_{AGI}(t)),$$

causing $\dot{K}_{AGI}(t) < 0$. Consequently, $K_{AGI}(t)$ cannot diverge and instead converges to a finite steady-state value K_{AGI}^{∞}. Given this finite steady-state capital stock, and since $L_h(t) \geq \bar{L}_h > 0$ by assumption, the production function $Y(t)$, which aggregates AGI capital and labor through a CES function, ensures that the marginal product of labor remains strictly positive

$$w_h(t) = \frac{\partial Y(t)}{\partial L_h(t)} \to w_h^{\infty} > 0.$$

Thus, human wages stabilize at a positive level, and human labor retains economic relevance. The long-run share of income accruing to humans (human economic power) remains strictly positive. Therefore, in the superlinear cost regime, the economy converges to a stable equilibrium that preserves human participation in production and distribution.

(ii) Sublinear Cost Case. Assume $V(K_{AGI})$ is sublinear, i.e.,

$$\lim_{K_{AGI}\to\infty} \frac{V(K_{AGI})}{K_{AGI}} = 0.$$

By Theorem 5, sublinear costs imply that for large K_{AGI}, the cost term $V(K_{AGI})$ grows too slowly to offset the linear and R&D-driven growth terms in the accumulation equation. Consequently,

$$\dot{K}_{AGI}(t) > 0 \quad \text{for large } K_{AGI},$$

which implies

$$\lim_{t\to\infty} K_{AGI}(t) = \infty.$$

As $K_{AGI}(t) \to \infty$, the CES production function implies that AGI capital increasingly dominates aggregate output. The marginal product of human labor, given by

$$w_h(t) = A\left(\delta_K(K + K_{AGI}(t))^{\rho} + \delta_L L_h(t)^{\rho}\right)^{\frac{1}{\rho}-1} \delta_L L_h(t)^{\rho-1},$$

declines toward zero, because the capital term dominates and the expression becomes insensitive to variations in $L_h(t)$. By Proposition 15, wages collapse

$$\lim_{t\to\infty} w_h(t) = 0,$$

and firms reduce labor demand to the technological lower bound

$$\lim_{t\to\infty} L_h(t) = \bar{L}_h.$$

With human labor earning zero wages and receiving no capital income (since $\tau = 0$ in this scenario), human economic power $P_h(t)$ collapses

$$\lim_{t\to\infty} P_h(t) = 0.$$

Therefore, the asymptotic outcome of the economy is fully determined by the curvature of $V(K_{AGI})$

- Superlinear costs $\Longrightarrow$ bounded K_{AGI}, positive wages, and survival of human economic power.
- Sublinear costs $\Longrightarrow$ unbounded K_{AGI}, wage collapse, and vanishing human economic power.

Thus, superlinear cost scaling is a *necessary and sufficient* structural condition for preserving human economic relevance in the absence of redistributive interventions. □

Theorem 6 establishes a sharp bifurcation in long-run economic and social outcomes, determined entirely by the asymptotic structure of AGI operating costs. When variable costs are *superlinear*, AGI capital accumulation is endogenously self-limiting: marginal costs rise faster than productive returns, causing expansion to stabilize. This bounded growth anchors AGI at a finite scale, ensuring that human labor retains economic relevance. In this regime, wages stabilize at positive levels, labor demand persists, and human agents continue to earn market-based income. The result is a technologically transformed, yet socially inclusive economy in which the fundamental tenets of the social contract—participation, reciprocity, and fair access to economic resources—are preserved without the need for external redistribution. By contrast, if variable costs are *sublinear*, AGI accumulation becomes unstable and potentially explosive. As AGI capital scales with negligible marginal costs, it increasingly displaces human labor in production. Wages collapse, employment contracts to the technological minimum, and market-based human income effectively disappears. This outcome severs the link between individual contribution and reward, hollowing out the social contract and generating extreme distributional inequality. Human beings, though biologically present, become economically marginalized and socially disenfranchised. Without corrective institutional mechanisms—such as redistribution, public ownership, or enforced scaling frictions—the economy risks devolving into a dualistic structure of AGI capital holders and economically irrelevant humans. Thus, the curvature of AGI cost functions is not merely a technological detail, but a structural determinant of long-run social inclusion. Superlinear costs function as a natural bulwark against economic exclusion, while sublinear costs necessitate active policy intervention to prevent the dissolution of economic citizenship and civic equality.

Theorem 7 (Transitional Shocks Due to Stepwise Fixed Costs)**.** *Suppose $F(t)$ adjusts discretely every 5 periods, such that*

$$F(t) = F_n \quad \textit{for} \quad t \in [5n, 5(n+1)), \quad n \in \mathbb{N}_0, \tag{47}$$

and $F_n < F_{n+1}$. Then, at each discrete transition point $t = 5n$, the growth rate $\dot{K}_{AGI}(t)$ experiences a downward discontinuity given by

$$\Delta \dot{K}_{AGI} = -(F_{n+1} - F_n) < 0. \tag{48}$$

This shock may cause

(i) A temporary switch from net accumulation to net decumulation of $K_{AGI}(t)$,

(ii) Transitional contraction in AGI-driven output and associated variables,

(iii) Eventual re-stabilization depending on the convexity of $V(K_{AGI})$ and aggregate demand.

Proof. At each transition point $t = 5n$, the fixed cost $F(t)$ increases discretely from F_n to F_{n+1}, with $F_{n+1} > F_n$ by assumption. The law of motion for AGI capital is

$$\dot{K}_{AGI}(t) = \phi s_R Y(t) + (\theta - \delta_{AGI})K_{AGI}(t) - F(t) - V(K_{AGI}(t)).$$

At the moment of transition, all terms except $F(t)$ are continuous in t. Therefore, the jump in $F(t)$ causes an instantaneous and discrete change in $\dot{K}_{AGI}(t)$ of magnitude

$$\Delta \dot{K}_{AGI} = -(F_{n+1} - F_n).$$

This discontinuity is strictly negative because $F_{n+1} > F_n$:

$$\Delta \dot{K}_{AGI} < 0.$$

Case 1: $\dot{K}_{AGI}(t^-)$ positive before the shock. If AGI capital was growing before the shock, the negative jump in $\dot{K}_{AGI}(t)$ may reduce the growth rate, or if large enough, bring it to zero or negative. In the latter case, the economy temporarily switches to net decumulation:

$$\dot{K}_{AGI}(t^+) < 0.$$

During this phase, $K_{AGI}(t)$ declines.

Case 2: Effect on Output and Further Feedback. A contraction in $K_{AGI}(t)$ reduces $Y(t)$, since $Y(t)$ is assumed to depend positively on $K_{AGI}(t)$. This lowers $\phi s_R Y(t)$, which is a source term in $\dot{K}_{AGI}(t)$, reinforcing the downward pressure on AGI accumulation.

Case 3: Eventual Stabilization. However, as $K_{AGI}(t)$ declines, the term $V(K_{AGI}(t))$ also declines due to monotonicity and convexity of $V(\cdot)$. Eventually, a new balance may be reached where

$$\phi s_R Y(t) + (\theta - \delta_{AGI})K_{AGI}(t) = F_{n+1} + V(K_{AGI}(t)),$$

at which point $\dot{K}_{AGI}(t) = 0$ and the system stabilizes at a new steady-state K_{AGI}^{n+1}. The likelihood and speed of this re-stabilization depend critically on

- the convexity of $V(\cdot)$ (which governs how quickly costs fall when $K_{AGI}(t)$ declines), and
- the responsiveness of $Y(t)$ to reductions in $K_{AGI}(t)$ (i.e., aggregate demand conditions).

Thus, each discrete increase in $F(t)$ induces a negative shock to $\dot{K}_{AGI}(t)$, potentially causing temporary decumulation of AGI capital and contraction in AGI-driven output. Over time, however, provided that $V(\cdot)$ is sufficiently convex and $Y(t)$ does not collapse, the system can transition to a new steady state with re-stabilized $K_{AGI}(t)$. □

Proposition 16 (Labor and Wage Impacts from Transitional Shocks). *Under the conditions of Theorem 7, suppose firms hire human labor $L_h(t)$ to maximize profits subject to a CES production function*

$$Y(t) = A\left(\delta_K(K + K_{AGI}(t))^\rho + \delta_L L_h(t)^\rho\right)^{1/\rho}, \quad \text{with } 0 < \rho < 1. \tag{49}$$

Let wages be determined by marginal productivity

$$\frac{\partial Y(t)}{\partial L_h(t)} = w_h(t). \tag{50}$$

Then a transitional cost shock at $t = 5n$ that reduces $\dot{K}_{AGI}(t)$ and slows AGI growth yields:

(i) A temporary increase in the marginal product of human labor,

(ii) A non-monotonic rise in wages $w_h(t)$,

(iii) A possible increase in $L_h(t)$ if wage increases do not fully offset the relative productivity gain from substituting toward labor.

Proof. The firm maximizes profit and hires labor according to the first-order condition:

$$w_h(t) = \frac{\partial Y(t)}{\partial L_h(t)} = A\left(\delta_K(K + K_{AGI}(t))^\rho + \delta_L L_h(t)^\rho\right)^{\frac{1}{\rho}-1} \cdot \delta_L \rho L_h(t)^{\rho-1}.$$

At time $t = 5n$, Theorem 7 implies a discrete upward jump in $F(t)$, causing an instantaneous reduction in $\dot{K}_{AGI}(t)$. As a result, the rate of accumulation of $K_{AGI}(t)$ slows, or $K_{AGI}(t)$ may temporarily decline.

(i) The term $(K + K_{AGI}(t))^\rho$ falls, reducing the effective capital contribution to the production function. Since CES production implies imperfect substitutability between labor and AGI capital, the marginal productivity of human labor increases

$$\frac{\partial Y(t)}{\partial L_h(t)} \uparrow \quad \text{as} \quad K_{AGI}(t) \downarrow .$$

(ii) As $w_h(t)$ equals marginal productivity, the rise in $\frac{\partial Y}{\partial L_h}$ implies a corresponding increase in wages. This increase may not persist if $K_{AGI}(t)$ resumes growth later, hence is typically non-monotonic.

(iii) If the increase in $w_h(t)$ is less than the increase in marginal product, the inequality $\frac{\partial Y}{\partial L_h} > w_h(t)$ holds, incentivizing firms to hire more human labor. Thus, $L_h(t)$ may temporarily increase in response to the capital shock. □

Theorems 6 and 7 together with Proposition 16 reveal the deep interplay between long-run structural forces and short-run transitional dynamics in shaping human labor outcomes within AGI-dominated economies. From a *structural* perspective, the curvature of the AGI cost function $V(K)$ plays a decisive role in determining long-run economic inclusion. Higher curvature (superlinear costs) limits the steady-state accumulation of AGI capital K_{AGI}^{∞}, thereby preserving a role for human labor in production. In such cases, human wages and employment stabilize at strictly positive levels, ensuring that labor retains an economically meaningful and socially dignified role in the economy. This structural brake on AGI expansion mitigates displacement and upholds a more equitable distribution of economic surplus across capital and labor. From a *transitional* perspective, stepwise increases in fixed costs $F(t)$—which may correspond to regulatory shifts, infrastructure scaling, or policy interventions—introduce short-run shocks that disrupt AGI accumulation. When these shocks occur, as shown in Proposition 16, the temporary reduction or reversal in AGI capital growth raises the marginal productivity of human labor, triggering non-monotonic increases in wages and potentially boosting employment. These episodic rebalancing effects restore, if only transiently, human labor's importance in production and income generation.

Taken together, these results emphasize that both long-run structural frictions and short-run transitional shocks are critical levers for safeguarding human economic relevance. The shape of $V(K)$ determines the asymptotic limits of automation, while adjustments to $F(t)$ shape cyclical patterns of inclusion and exclusion. Critically, while transitional shocks generate socially meaningful windows during which labor regains power and income, they are inherently temporary. Thus, relying on them alone is inadequate for upholding the deeper normative commitments of the social contract—namely, inclusion, reciprocity, and shared prosperity. Accordingly, a robust social contract in AGI-driven economies will likely require both endogenous technological frictions (via superlinear cost structures) and deliberate institutional interventions (such as redistribution, wage supports, or labor market regulation). Together, these mechanisms can stabilize labor's role not merely episodically, but durably and equitably, ensuring that technological progress remains aligned with human flourishing.

Theorem 8 (Sensitivity of AGI Stability to Cost Curvature)**.** *Let two economies be characterized by variable cost functions $V_1(K)$ and $V_2(K)$ such that*

$$V_1''(K) > V_2''(K) \geq 0 \quad \textit{for all sufficiently large } K. \tag{51}$$

Assume both economies share the same parameters $(\theta, \delta_{AGI}, \phi, s_R)$ and long-run fixed cost F^∞. Let $Y^\infty(K)$ be strictly increasing in K. Then the corresponding steady-state capital levels satisfy

$$K_{AGI}^{(1),\infty} < K_{AGI}^{(2),\infty}. \tag{52}$$

Proof. Define for each economy $i \in \{1, 2\}$ the steady-state equation

$$G(K) := (\theta - \delta_{AGI})K + \phi s_R Y^\infty(K) = F^\infty + V_i(K) =: H_i(K),$$

where $Y^\infty(K)$ is assumed to be continuously differentiable and strictly increasing, and $V_i(K)$ is twice continuously differentiable with $V_i'(K) > 0$ and $V_i''(K) \geq 0$. Let us now analyze the properties of both sides: $G(K)$ is strictly increasing in K, since both $K \mapsto (\theta - \delta_{AGI})K$ and $K \mapsto Y^\infty(K)$ are strictly increasing. $H_i(K) = F^\infty + V_i(K)$ is strictly increasing in K, as $V_i'(K) > 0$. For large K, since $V_1''(K) > V_2''(K)$, the rate of increase of $H_1(K)$ exceeds that of $H_2(K)$. Now consider the difference function $D(K) := H_1(K) - H_2(K) = V_1(K) - V_2(K)$. Since $V_1''(K) > V_2''(K)$ for sufficiently large K, we have

$$D'(K) = V_1'(K) - V_2'(K) > 0 \quad \text{for large } K.$$

Thus, $D(K)$ is strictly increasing on that interval, meaning the gap between $H_1(K)$ and $H_2(K)$ widens with increasing K. Because both $G(K)$ and $H_i(K)$ are strictly increasing and continuous, each steady-state equation $G(K) = H_i(K)$ has a unique solution by the intermediate value theorem. Suppose $K^{(1),\infty}$ and $K^{(2),\infty}$ solve the steady-state equations for economies 1 and 2, respectively. Since $H_1(K) > H_2(K)$ for all sufficiently large K, and the left-hand side $G(K)$ is common, it must intersect the larger of the two curves later. Thus

$$G(K^{(1),\infty}) = H_1(K^{(1),\infty}) < H_2(K^{(1),\infty}) < H_2(K^{(2),\infty}) = G(K^{(2),\infty}),$$

and since G is strictly increasing, it follows that

$$K^{(1),\infty} < K^{(2),\infty}.$$

Therefore, higher cost curvature implies a lower steady-state level of AGI capital. □

Proposition 17 (Cost Curvature and Human Labor Outcomes)**.** *Under the assumptions of Theorem 8, suppose that output $Y^\infty(K, L_h)$ is produced via a CES function where AGI capital and human labor are imperfect substitutes. Then* *(i) In the economy with higher cost curvature $V_1(K)$, the steady-state human wage w_h^∞ and labor employment L_h^∞ are both strictly higher than in the economy with lower cost curvature $V_2(K)$.*

(ii) Consequently, the long-run human income share is strictly larger under V_1 than under V_2.

Proof. From Theorem 8, we know that

$$K_{AGI}^{(1),\infty} < K_{AGI}^{(2),\infty},$$

where $K_{AGI}^{(i),\infty}$ denotes the steady-state AGI capital in economy $i \in \{1, 2\}$. Assume output is generated via a CES production function

$$Y(K, L_h) = A\left(\delta_K K^\rho + \delta_L L_h^\rho\right)^{\frac{1}{\rho}},$$

where $\rho \in (0, 1)$ controls the elasticity of substitution between AGI capital and human labor. The marginal product of human labor is

$$\frac{\partial Y}{\partial L_h} = A\left(\delta_K K^\rho + \delta_L L_h^\rho\right)^{\frac{1}{\rho}-1} \delta_L L_h^{\rho-1}.$$

Since $\frac{1}{\rho} - 1 > 0$, the term $\left(\delta_K K^\rho + \delta_L L_h^\rho\right)^{\frac{1}{\rho}-1}$ is decreasing in K for fixed L_h, because the elasticity parameter $\rho < 1$ implies diminishing returns to each input. Now, observe

- In economy 1, $K = K_{AGI}^{(1),\infty}$ is lower than in economy 2, where $K = K_{AGI}^{(2),\infty}$.
- Therefore, the marginal product of labor satisfies

$$\left.\frac{\partial Y}{\partial L_h}\right|_{K=K_{AGI}^{(1),\infty}} > \left.\frac{\partial Y}{\partial L_h}\right|_{K=K_{AGI}^{(2),\infty}},$$

  holding L_h constant.
- Under firm profit maximization, the wage satisfies

$$w_h = \frac{\partial Y}{\partial L_h}.$$

  Hence,

$$w_h^{(1),\infty} > w_h^{(2),\infty}.$$

- Furthermore, higher wages imply that firms have an incentive to maintain a higher level of human employment, assuming a standard downward-sloping labor demand curve.

Therefore, both the long-run human wage w_h and employment level L_h are higher in the economy with greater cost convexity (i.e., economy 1). □

Proposition 18 (Cost Curvature and Human Labor Share)**.** *Assume the conditions of Theorem 8 hold. Further suppose the production function $Y(K_{AGI}, L_h)$ is CES with elasticity of substitution $\sigma \in (0, \infty)$ and that firms hire human labor according to*

$$\frac{\partial Y}{\partial L_h} = w_h. \tag{53}$$

Then the economy with higher cost curvature $V_1(K)$ satisfies

(i) $w_h^{(1),\infty} > w_h^{(2),\infty}$,

(ii) $L_h^{(1),\infty} > L_h^{(2),\infty}$,

(iii) The long-run labor share of income is higher under V_1 than under V_2.

Proof. By Theorem 8, we have $K_{AGI}^{(1),\infty} < K_{AGI}^{(2),\infty}$. In a CES production function,

$$Y(K, L_h) = A\left(\delta_K K^\rho + \delta_L L_h^\rho\right)^{1/\rho}, \quad \text{with } \rho = \frac{\sigma - 1}{\sigma},$$

and the marginal product of labor is

$$\frac{\partial Y}{\partial L_h} = A\left(\delta_K K^\rho + \delta_L L_h^\rho\right)^{\frac{1}{\rho}-1} \delta_L L_h^{\rho-1}.$$

Holding L_h constant, this expression is decreasing in K. Thus, lower $K_{AGI}^{(1),\infty}$ implies a higher $\partial Y/\partial L_h$ and hence higher w_h in equilibrium

$$w_h^{(1),\infty} > w_h^{(2),\infty}.$$

Higher wages, under standard downward-sloping demand for labor, also imply

$$L_h^{(1),\infty} > L_h^{(2),\infty},$$

and hence a larger total labor income $w_h L_h$. Given total income is determined by $Y(K_{AGI}, L_h)$, the labor share $w_h L_h / Y$ is also larger in the high-curvature economy. □

Theorem 8 and Propositions 17–18 jointly illustrate how the curvature of AGI variable costs fundamentally shapes both macroeconomic equilibrium and distributive outcomes in AGI-driven economies. From a macroeconomic perspective, higher cost curvature directly limits AGI capital accumulation by ensuring that operating costs grow disproportionately at large scales. This superlinear growth in costs acts as a structural check on runaway automation, yielding a lower steady-state AGI capital stock K_{AGI}^{∞}. In contrast, lower curvature permits AGI capital to expand further before marginal costs become prohibitive, resulting in higher equilibrium levels of automation. Thus, curvature governs the *degree of AGI saturation* in the long-run production structure. From a distributional perspective, this asymptotic limitation on AGI capital accumulation has profound implications for human labor. Because the CES production function implies imperfect substitutability between labor and capital, higher steady-state AGI capital suppresses the marginal product of human labor. As shown in Propositions 17 and 18, economies with higher cost curvature—and thus lower K_{AGI}^{∞}—feature higher marginal productivity of labor, leading to:

- higher steady-state wages w_h^{∞},
- higher human employment levels L_h^{∞}, and
- a larger long-run labor share of income.

Viewed through the lens of the social contract, these results highlight that technological cost structures are not merely technical parameters but fundamental determinants of economic inclusion and human relevance. When cost curvature is high, labor markets retain strength, human workers earn higher wages, and the surplus generated by AGI-enhanced production is shared more broadly. When curvature is low, automation expands unchecked, eroding human wages, shrinking employment, and concentrating income in AGI capital holders. Thus, the curvature of $V(K)$ emerges as a central policy-relevant lever. Technological design choices (e.g., architectures with higher coordination costs), regulatory interventions (e.g., taxation of scale), or deliberate imposition of artificial scaling frictions can all serve to increase cost curvature, thereby preserving human labor's role in economic life. In sum, these results collectively reveal that *superlinear cost structures are not merely stabilizing forces for AGI dynamics, but essential safeguards for distributive justice and social stability in automated economies.*

5 Distributional Bifurcations and Human Welfare Regimes

As AGI capital increasingly dominates economic production, the distribution of income between human labor and AGI systems emerges as a key determinant of macroeconomic stability and long-run welfare. This section extends the model to endogenize aggregate demand, highlighting how divergent consumption propensities between labor earners and capital owners generate feedback loops that influence economic dynamics. In particular, as income shifts from labor to capital, aggregate demand may contract, constraining output and inducing bifurcations between stagnation and inclusive growth. The analysis formalizes how redistributive institutions influence not only equity, but also the macroeconomic regime toward which the economy ultimately converges. In the extended model, suppose that aggregate demand $D(t)$ is a function of the income distribution between AGI capital owners and human labor, i.e.,

$$D(t) = c_h Y_h(t) + c_K Y_K(t), \tag{54}$$

where $c_h > c_K$ reflect differing marginal propensities to consume. Human income $Y_h(t)$ consists of wage earnings $w_h(t)L_h(t)$, while $Y_K(t)$ captures AGI and capital returns.

Assumption 9 (Demand Feedback and Consumption Concentration)**.** *If AGI capital dominates total output while human labor income collapses, then $D(t)$ stagnates or contracts due to low c_K.*

The assumption reflects that AGI owners typically exhibit lower marginal propensities to consume (c_K) than wage earners, whose consumption depends strongly on labor income. As AGI capital dominates production and wages collapse, aggregate demand becomes increasingly dependent on AGI capital income, which does not translate proportionally into consumption. This causes $D(t)$ to stagnate or contract despite rising output. This mechanism is widely recognized in macroeconomic theory (see e.g. Mian et al. (2021)), where rising income concentration suppresses demand and can destabilize growth dynamics.

Theorem 10 (Bifurcation in Steady-State Outcomes)**.** *Suppose AGI capital accumulation is governed by*

$$\dot{K}_{AGI}(t) = \phi s_R \min\{Y(t), D(t)\} + (\theta - \delta_{AGI})K_{AGI}(t) - F^{\infty} - V(K_{AGI}(t)), \tag{55}$$

and that output $Y(t)$ depends positively on $K_{AGI}(t)$ and $L_h(t)$, while aggregate demand is given by

$$D(t) = c_h Y_h(t) + c_K Y_K(t), \tag{56}$$

where $Y_h(t) = w_h(t)L_h(t)$, $Y_K(t) = r_K K + r_{AGI} K_{AGI}(t)$, and $c_h > c_K$. Then if $Y_h(t) \to 0$ while $Y_K(t) \to \infty$, there exists a threshold beyond which:

(i) The economy converges to a high-AGI, low-human-income equilibrium where $\min\{Y(t), D(t)\} = D(t)$ is constrained,

(ii) A distinct equilibrium with bounded K_{AGI}^{∞} and sustained human income exists, supported by stronger demand and redistribution.

Proof. The key feature of the model is that AGI capital accumulation depends on the constrained source of R&D investment, which is given by

$$\dot{K}_{AGI}(t) = \phi s_R \min\{Y(t), D(t)\} + (\theta - \delta_{AGI})K_{AGI}(t) - F^{\infty} - V(K_{AGI}(t)).$$

Thus, the growth of $K_{AGI}(t)$ is demand-constrained if $D(t) < Y(t)$, and output-constrained if $Y(t) < D(t)$. The dynamic bifurcation arises from how $D(t)$ evolves with increasing AGI capital and falling human income.

Step 1: Analyze demand under rising AGI dominance. By assumption,

$$D(t) = c_h Y_h(t) + c_K Y_K(t),$$

where

$$Y_h(t) = w_h(t)L_h(t) \quad \text{and} \quad Y_K(t) = r_K K + r_{AGI} K_{AGI}(t).$$

As $K_{AGI}(t)$ grows large and human employment $L_h(t)$ contracts (or $w_h(t) \to 0$), it follows that $Y_K(t)$ grows large, because $K_{AGI}(t) \to \infty$, and $Y_h(t)$ shrinks or vanishes, because $w_h(t)L_h(t) \to 0$. Thus, asymptotically,

$$D(t) \approx c_K Y_K(t).$$

But since $c_K < c_h$ by assumption, demand grows more slowly than output $Y(t)$, which depends directly on $K_{AGI}(t)$ and grows faster. Therefore,

$$\lim_{t\to\infty} \frac{D(t)}{Y(t)} < 1.$$

Eventually, this implies that

$$\min\{Y(t), D(t)\} = D(t),$$

i.e., demand becomes the *binding constraint* on AGI accumulation.

Step 2: Consequences of demand-constrained accumulation. Once $D(t)$ binds, the AGI capital law of motion becomes

$$\dot{K}_{AGI}(t) = \phi s_R D(t) + (\theta - \delta_{AGI})K_{AGI}(t) - F^{\infty} - V(K_{AGI}(t)).$$

In the limit, $D(t) \approx c_K Y_K(t) = c_K r_{AGI} K_{AGI}(t)$, so

$$\dot{K}_{AGI}(t) = [\phi s_R c_K r_{AGI} + (\theta - \delta_{AGI})] K_{AGI}(t) - F^{\infty} - V(K_{AGI}(t)).$$

Case A: If $V(K_{AGI})$ is sufficiently convex (superlinear), then for large K_{AGI} the term $V(K_{AGI})$ will eventually dominate the linear accumulation terms. Hence, $\dot{K}_{AGI}(t)$ will become negative at high K_{AGI}, and by continuity and the intermediate value theorem, there will exist a finite K_{AGI}^{∞} such that

$$\dot{K}_{AGI}(t) = 0.$$

This gives a steady state with large AGI capital K_{AGI}^{∞}, very small or near-zero human labor income Y_h^{∞}, low consumption demand due to low c_K, and stagnant AGI growth — the high-AGI stagnation regime.

Step 3: Alternative equilibrium with stronger demand.

Case B: If $Y_h(t)$ remains substantial due to redistribution or policy (e.g. universal basic income, labor market interventions), then $D(t)$ increases because $c_h > c_K$, so demand grows faster, $\min\{Y(t), D(t)\} = Y(t)$ because $D(t)$ no longer binds, and AGI accumulation remains output-driven rather than demand-driven. In this case, the steady state satisfies

$$\dot{K}_{AGI}(t) = \phi s_R Y(t) + (\theta - \delta_{AGI})K_{AGI}(t) - F^{\infty} - V(K_{AGI}(t)) = 0.$$

The higher $Y(t)$ raises the feasible steady-state K_{AGI}^{∞} and sustains positive wages and labor income. Thus, we obtain a distinct equilibrium characterized by bounded but moderate AGI capital, meaningful human wages and employment, higher aggregate demand, and dynamic rather than stagnant AGI accumulation — the moderate-AGI equitable regime. Therefore, depending on the level of human labor income and aggregate demand dynamics, the economy bifurcates into

1. a demand-constrained, stagnation-prone equilibrium dominated by AGI capital and low human relevance,
2. and a higher-demand, more inclusive equilibrium sustaining human wages and stabilizing AGI accumulation.

The existence of these two distinct regimes completes the proof. □

Theorem 10 reveals a fundamental bifurcation in the long-run trajectories of AGI economies, driven by the endogenous interaction between income distribution and aggregate demand. When AGI capital dominates production and human wages collapse, the low marginal propensity to consume of capital owners depresses aggregate demand. This demand contraction becomes a self-reinforcing brake on further AGI accumulation, producing a stagnation equilibrium characterized by high AGI capital, low output growth, and extreme human economic exclusion. In this regime, humans are economically marginalized not merely by technological obsolescence but by the systemic insufficiency of demand, which suppresses productive investment and curtails further growth. Conversely, when policy or institutional interventions ensure that human income remains sufficiently robust—through redistribution, labor policy, or public transfers—aggregate demand rises relative to AGI-centric output. This prevents demand from becoming the binding constraint, enabling a dynamic equilibrium in which AGI accumulation continues in tandem with sustained human labor participation and wages. Such an equilibrium preserves macroeconomic dynamism while maintaining inclusive income distribution. The bifurcation thus carries profound implications for the social contract. In the absence of redistribution, the economy naturally tends toward a high-AGI stagnation trap, where humans are deprived of both economic agency and participation in societal surplus. By contrast, deliberate policy action can shift the economy into a moderate-AGI equilibrium that sustains human welfare and preserves the normative ideals of reciprocity, inclusion, and dignity in economic life. The model therefore formalizes the critical role of distributive institutions in preventing technological progress from undermining the moral foundations of society.

Let $U(t)$ denote a social welfare function aggregating labor income and employment given by

$$U(t) = u(w_h(t), L_h(t)) = \log(w_h(t)) + \alpha \log(L_h(t)), \quad \alpha > 0.$$

Then the steady-state value U^{∞} differs across equilibria. In the high-concentration regime, $w_h \to 0$ and $L_h \to 0$, so $U^{\infty} \to -\infty$. In the moderate-AGI regime, $w_h^{\infty} > 0$ and $L_h^{\infty} > 0$, so U^{∞} is finite and higher.

Corollary 1 (Welfare Gap across Regimes). *Let U_1^∞ and U_2^∞ denote the long-run average human welfare under two regimes*

- *Regime 1: Moderate AGI accumulation, positive human wages $w_h^\infty > 0$ and employment $L_h^\infty > 0$,*
- *Regime 2: Excessive AGI accumulation, collapse of human wages and labor $w_h^\infty = 0$, $L_h^\infty = 0$.*

Assume individual human welfare is increasing in labor income and transfer income. Then

$$U_1^\infty > U_2^\infty, \tag{57}$$

whenever $w_h^\infty, L_h^\infty > 0$ under Regime 1.

Proof. Define human social welfare at time t by the function

$$U(t) = u(w_h(t), L_h(t)) = \log(w_h(t)) + \alpha \log(L_h(t)),$$

where $\alpha > 0$ reflects the weight placed on employment in social welfare. By assumption, $U(t)$ is strictly increasing in both human wages $w_h(t)$ and employment $L_h(t)$. Consider the steady-state values U_1^∞ and U_2^∞ under Regimes 1 and 2, respectively. In Regime 1, by definition, the steady-state human wage w_h^∞ and labor employment L_h^∞ are strictly positive. Therefore, the arguments of the logarithmic terms in $U(t)$ remain strictly positive in the steady state, and both $\log(w_h^\infty)$ and $\log(L_h^\infty)$ are finite real numbers. As a result, the steady-state welfare U_1^∞ is finite and well-defined. In contrast, under Regime 2, excessive AGI accumulation drives both human wages and employment to zero asymptotically, so that

$$\lim_{t\to\infty} w_h(t) = 0 \quad \text{and} \quad \lim_{t\to\infty} L_h(t) = 0.$$

Given the properties of the logarithmic function, $\log(0)$ is undefined in the real numbers and approaches $-\infty$ as its argument approaches zero from above. Therefore, as $w_h(t) \to 0$ and $L_h(t) \to 0$, the welfare function $U(t)$ diverges negatively

$$\lim_{t\to\infty} U(t) = -\infty.$$

This establishes that $U_2^\infty = -\infty$ in the high-concentration regime. To complete the argument, observe that any finite value U_1^∞, no matter how small, strictly exceeds $U_2^\infty = -\infty$. Thus,

$$U_1^\infty > U_2^\infty.$$

This inequality holds as long as $w_h^\infty > 0$ and $L_h^\infty > 0$ under Regime 1, which is true by assumption. Hence, the economy that stabilizes in the moderate AGI regime, with continued human employment and positive wages, guarantees strictly higher long-run social welfare compared to the economy that transitions into the high-concentration regime where human labor is economically irrelevant. This completes the proof. □

Corollary 1 reveals a bifurcation in human welfare trajectories. In the moderate AGI regime, where cost frictions constrain AGI expansion, labor remains economically relevant. Positive wages and employment sustain aggregate demand and ensure that social welfare remains finite, upholding the market's role in rewarding participation. By contrast, in the high-concentration regime, runaway AGI accumulation drives wages and employment to zero, rendering human labor economically obsolete. The resulting collapse in labor income leads welfare to diverge negatively, even under redistributive schemes, as transfers rarely fully substitute lost earned income. In this regime, economic reciprocity is severed and the social contract is effectively broken. Thus, the model highlights that regulating AGI scalability and ensuring inclusive income distribution are not only economic challenges, but also essential to preserving social cohesion.

Theorem 11 (Existence of a Rational Social Contract under AGI Expansion). *Suppose*

(i) *Human welfare $U(t)$ depends positively on disposable income: $U(t) = u(w_h(t)L_h(t) + T(t))$, where $u(\cdot)$ is strictly increasing and concave,*

(ii) *AGI expansion leads to endogenous bifurcation between a low-human-welfare regime (collapse of $w_h(t)$, $L_h(t)$) and a positive-human-welfare regime (sustained $w_h(t), L_h(t) > 0$),* *(iii) Humans collectively possess the ability to enforce redistribution mechanisms (e.g., through taxation τ or operational constraints on AGI).*

Then there exists a redistribution policy $\tau^\star \in (0,1)$ and regulatory regime such that

$$\lim_{t\to\infty} U^{regulated}(t) > \lim_{t\to\infty} U^{unregulated}(t), \tag{58}$$

where $U^{regulated}(t)$ denotes welfare under an enforced Social Contract. Thus, a rational new Social Contract exists, where society imposes redistribution or AGI constraints to prevent welfare collapse.

Proof. Without redistribution ($\tau = 0$), by assumption (ii), as $t \to \infty$

$$w_h(t)L_h(t) \to 0, \quad T(t) = 0,$$

hence

$$\lim_{t\to\infty} U^{\text{unregulated}}(t) = u(0).$$

Since u is strictly increasing, $u(0)$ is the minimum attainable utility level (often interpreted as the utility of extreme poverty). Now, suppose we implement redistribution at a constant rate $\tau > 0$. Then by assumption (iii), the transfer satisfies

$$T(t) = \tau r_{AGI} K_{AGI}(t),$$

and as $K_{AGI}(t) \to \infty$ (under unbounded AGI growth), we have

$$T(t) \to +\infty \quad \text{as} \quad t \to \infty.$$

Thus, human disposable income $w_h(t)L_h(t) + T(t)$ eventually becomes dominated by $T(t)$ and grows unboundedly. Since u is strictly increasing but strictly concave, $\lim_{x\to\infty} u(x)$ is finite but larger than $u(0)$. That is

$$\lim_{t\to\infty} U^{\text{regulated}}(t) = \lim_{x\to\infty} u(x) > u(0).$$

Hence,

$$\lim_{t\to\infty} U^{\text{regulated}}(t) > \lim_{t\to\infty} U^{\text{unregulated}}(t).$$

Therefore, there exists some $\tau^\star > 0$ (not necessarily unique) such that enforcing redistribution at rate $\tau^\star$ ensures humans attain strictly higher long-run welfare compared to no redistribution. □

Theorem 11 provides a formal justification for the re-emergence of the social contract as a necessary economic institution in the face of AGI-driven technological bifurcation. When left unregulated, AGI expansion drives labor wages and employment toward collapse, pushing human welfare toward subsistence levels and eroding the economic reciprocity essential to social cohesion. However, the theorem shows that societies capable of enforcing redistribution or AGI constraints can rationally avoid this dystopian outcome. By instituting transfer mechanisms funded through AGI rents, collective action can raise human welfare above subsistence, ensuring positive and stable utility levels even when market-based labor income vanishes. Crucially, this result is not normative but instrumental: it follows directly from rational welfare maximization under concave utility. Thus, the social contract reasserts itself not merely as a moral imperative, but as an economically optimal institutional response to prevent collapse of mass welfare in an increasingly AGI-dominated economy. This reframing suggests that redistributive and regulatory regimes are endogenous and efficiency-enhancing outcomes in advanced technological societies, rather than external impositions or distortions of the market order.

Theorem 12 (Existence of Stationary Redistribution Agreements)**.** *Consider a dynamic game between human agents and AGI capital owners, where at each period agents negotiate a redistribution rate $\tau(t) \in [0,1]$. Assume*

(i) Agents maximize discounted utilities $U_h = \sum_{t=0}^{\infty} \beta_h^t u_h(c_h(t))$ with $\beta_h \in (0,1)$,

(ii) AGI owners maximize discounted returns $U_{AGI} = \sum_{t=0}^{\infty} \beta_{AGI}^t u_{AGI}(c_{AGI}(t))$,

(iii) Aggregate production $Y(t)$ depends continuously and strictly monotonically on $K_{AGI}(t)$ and $L_h(t)$,

(iv) Redistribution $\tau(t)$ reduces AGI owners' incentives for capital accumulation,

(v) Payoffs and strategies are continuous and the state space is compact.

Then, there exists a Markov Perfect Equilibrium with a stationary redistribution rate $\tau^\star \in [0, 1]$ satisfying dynamic incentive compatibility.

Proof. We provide the proof in 5 steps.

Step 1: Define state and strategy spaces. At time t, the state variable is the AGI capital stock $K_{AGI}(t) \in [0, \bar{K}]$, where $\bar{K} < \infty$ ensures compactness of the state space. At each period, agents choose a redistribution rate $\tau(t) \in [0, 1]$, forming a compact and convex action space $A = [0, 1]$.

Step 2: Define per-period payoffs. Given state $K_{AGI}(t)$ and choice $\tau(t)$, the per-period utilities are

$$u_h(\tau(t), K_{AGI}(t)) = u_h\big(w_h(t)L_h(t) + \tau(t)r_{AGI}K_{AGI}(t)\big),$$
$$u_{AGI}(\tau(t), K_{AGI}(t)) = u_{AGI}\big((1 - \tau(t))r_{AGI}K_{AGI}(t)\big).$$

By Assumptions (i)–(iv), these are continuous and strictly concave in $\tau(t)$.

Step 3: Dynamic programming formulation. The Bellman equations are

$$V_h(K) = \max_{\tau \in [0,1]} \Big\{u_h(\tau, K) + \beta_h \mathbb{E}[V_h(K')]\Big\},$$
$$V_{AGI}(K) = \max_{\tau \in [0,1]} \Big\{u_{AGI}(\tau, K) + \beta_{AGI} \mathbb{E}[V_{AGI}(K')]\Big\},$$

where K' is the next-period AGI capital stock, which depends continuously on K and τ.

Step 4: Existence of Markov Perfect Equilibrium. The game is a stochastic game with compact state and action spaces, continuous payoffs, and continuous transition dynamics. By standard results (see Ericson and Pakes (1995), or Maskin and Tirole (2001)), such a game admits at least one stationary Markov Perfect Equilibrium. Thus, there exists a stationary redistribution rate $\tau^\star$ satisfying

$$\tau(t) = \tau^\star \quad \text{for all } t \geq 0.$$

Step 5: Dynamic incentive compatibility. In equilibrium, $\tau^\star$ balances human welfare and AGI owners' investment incentives. If τ is too high, AGI owners underinvest, reducing future output. If too low, humans receive insufficient redistribution. Thus, $\tau^\star$ arises endogenously as a dynamically stable compromise where neither party has an incentive to deviate.

A stationary Markov Perfect Equilibrium with redistribution rate $\tau^\star \in [0, 1]$ exists and is dynamically incentive compatible. □

Theorem 12 extends Rousseau's classical idea of the Social Contract to a dynamic economy with AGI entities. Whereas the traditional Social Contract focused on balancing freedoms and protections among human citizens, the modern setting must account for non-human productive agents (AGI capital) whose incentives fundamentally influence societal outcomes. The stationary redistribution agreement $\tau^\star$ acts as a *dynamic social contract*: it continually balances human welfare against the productive incentives of AGI capital owners. Importantly, it guarantees existence even when agents' objectives are dynamically interdependent and potentially conflicting. Thus, this model generalizes Rousseau by formalizing incentive-compatible redistributive structures that stabilize both economic growth and social cohesion in a post-human-labor economy.

Theorem 13 (Political Stability Bifurcation)**.** *Let κ denote a measure of political power concentration among AGI owners. Then*

(i) If $\kappa < \kappa^\star$, there exists a stable positive redistribution $\tau^\star > 0$,

(ii) If $\kappa > \kappa^\star$, stable redistribution collapses and $\tau^\star = 0$ in equilibrium.

Thus, increasing inequality beyond $\kappa^\star$ endogenously destabilizes the social contract.

Proof. Let the redistribution rate τ be determined each period by a weighted political bargaining game, where the AGI owners and human agents have bargaining weights κ and $1-\kappa$, respectively. Let $U_h(\tau)$ and $U_{AGI}(\tau)$ be the indirect utility functions of the human population and AGI owners, respectively, with both functions assumed to be continuous and concave in τ. Define the Nash bargaining solution $\tau^\star(\kappa)$ as the solution to the following program

$$\tau^\star(\kappa) = \arg\max_{\tau\in[0,1]} \left[U_h(\tau)^{1-\kappa} \cdot U_{AGI}(\tau)^{\kappa}\right].$$

The first-order condition for an interior solution $\tau^\star > 0$ is

$$(1-\kappa)\frac{U_h'(\tau)}{U_h(\tau)} = \kappa\frac{U_{AGI}'(\tau)}{U_{AGI}(\tau)}.$$

As $\kappa \to 1$, the left-hand side approaches 0, while the right-hand side remains positive and increasing in τ due to AGI owners' aversion to redistribution. Therefore, the equality cannot hold for any $\tau > 0$ if κ is sufficiently large. Define the threshold $\kappa^\star$ as the supremum value such that the first-order condition still admits a positive solution. Then

(i) If $\kappa < \kappa^\star$, then there exists $\tau^\star > 0$ solving the bargaining condition.

(ii) If $\kappa > \kappa^\star$, then the optimal solution is at the boundary, $\tau^\star = 0$.

Hence, there is a bifurcation in the equilibrium redistribution rate $\tau^\star$ as κ crosses $\kappa^\star$. □

Theorem 13 provides a dynamic political economy foundation for the erosion of the social contract. As political power becomes too concentrated among AGI capital owners, their control over redistribution mechanisms ensures that no positive redistribution persists in equilibrium. This captures a critical rupture: even if positive redistribution would be efficient or desirable, excessive concentration of influence makes it politically unsustainable. In classical terms, Rousseau's Social Contract relies on mutual interdependence and reciprocal authority among citizens. In contrast, this result formalizes a breakdown of reciprocal cooperation: if one group (AGI owners) accumulates excessive structural dominance, then the social contract collapses—not from external shocks, but from internal political asymmetries. Thus, the theorem reinforces the necessity of institutional designs that check excessive concentration of influence, preserving a balanced polity where redistribution agreements can remain dynamically stable.

Theorem 14 (Fragility of Static Social Contracts under Technological Escalation)**.** *Suppose an initial redistribution contract $\tau(0)$ is agreed at $t = 0$ based on expected AGI capital dynamics. Let the actual capital accumulation be governed by*

$$\dot{K}_{AGI}(t) = \phi s_R Y(t) + (\theta - \delta_{AGI})K_{AGI}(t) - F^\infty - V(K_{AGI}(t)),$$

where $V(K)$ is continuous and convex, and $Y(t)$ depends on $K_{AGI}(t)$ and $L_h(t)$. If the realized growth rate of AGI capital exceeds a critical threshold $\theta^\dagger$, then there exists a finite $t^\dagger$ such that

$$\tau(t) \neq \tau(0) \quad \textit{for some } t > t^\dagger. \tag{59}$$

That is, the static redistribution agreement becomes dynamically unsustainable beyond $t^\dagger$.

Proof. At $t = 0$, suppose $\tau(0)$ is chosen based on expected capital growth $\hat{\theta} < \theta^\dagger$, implying an anticipated trajectory $K_{AGI}^{exp}(t)$ and revenue stream $T_{exp}(t) = \tau(0) r_{AGI} K_{AGI}^{exp}(t)$. Suppose the realized $\theta > \theta^\dagger$, so that $K_{AGI}(t)$ grows super-exponentially relative to $K_{AGI}^{exp}(t)$. Then actual transfers $T_{act}(t) = \tau(0) r_{AGI} K_{AGI}(t)$ may grow rapidly, leading to:

1. Incentive breakdown: AGI owners face diminished post-tax returns, reducing investment incentives.
2. Political backlash: AGI owners demand renegotiation, especially if $\tau(0)$ exceeds their preferred rate at high $K_{AGI}(t)$.
3. Strategic deviation: If either party (e.g., capital owners or state institutions) deviates from enforcing $\tau(0)$, then $\tau(t)$ must change.

These dynamics yield a contradiction with the assumption that $\tau(t) = \tau(0)$ for all t, implying the existence of $t^\dagger$ where the contract fails. Hence, fixed redistribution policies are not robust to unexpected surges in AGI productivity beyond a technological threshold $\theta^\dagger$. □

Theorem 14 formalizes a key vulnerability in classical social contracts: their static nature. Rousseau's ideal of mutual consent under common interest breaks down when technological capabilities diverge too rapidly between groups. Here, redistribution policies agreed under a shared understanding of future growth may collapse if AGI capital accelerates unexpectedly. This highlights a limitation of traditional contractarianism: it assumes bounded technological stability. In contrast, modern economies driven by self-improving AGI must incorporate dynamic renegotiation mechanisms—adaptive institutions that can recalibrate redistribution in response to endogenous shocks. Thus, a durable social contract under AGI requires not static agreement, but flexible governance anchored in dynamic incentive compatibility.

Corollary 2 (Welfare Gap across Regimes)**.** *Let U_1^∞ and U_2^∞ denote the long-run average human welfare under two regimes:*

- *Regime 1 (Moderate AGI accumulation): Human wages and employment converge to strictly positive steady-state levels, i.e., $w_h^\infty > 0$ and $L_h^\infty > 0$.*
- *Regime 2 (Excessive AGI accumulation): Human wages and employment collapse to zero, i.e., $w_h^\infty = 0$ and $L_h^\infty = 0$.*

Assume that human welfare at time t is given by

$$U(t) = u(w_h(t)L_h(t) + T(t)),$$

where $u(\cdot)$ is strictly increasing, and $T(t)$ is transfer income (possibly zero). Then

$$U_1^\infty > U_2^\infty,$$

whenever $w_h^\infty > 0$ and $L_h^\infty > 0$ under Regime 1.

Proof. By assumption, the utility function $u(\cdot)$ is strictly increasing. Therefore, to prove that $U_1^\infty > U_2^\infty$, it suffices to show that disposable income in Regime 1 exceeds that in Regime 2.

Regime 1. Since $w_h^\infty > 0$ and $L_h^\infty > 0$, labor income in steady state satisfies

$$w_h^\infty L_h^\infty > 0.$$

Hence, total steady-state disposable income is

$$Y_h^\infty = w_h^\infty L_h^\infty + T_1^\infty,$$

where T_1^∞ denotes transfer income in Regime 1. Since $w_h^\infty L_h^\infty > 0$, we immediately have

$$Y_h^\infty > T_1^\infty.$$

Regime 2. By assumption, $w_h^\infty = 0$ and $L_h^\infty = 0$, so labor income is zero. Thus, disposable income consists only of transfers

$$Y_h^{\infty,(2)} = T_2^\infty.$$

Comparison. Two cases arise: If $T_1^\infty \geq T_2^\infty$, then clearly

$$Y_h^\infty = w_h^\infty L_h^\infty + T_1^\infty > T_1^\infty \geq T_2^\infty = Y_h^{\infty,(2)}.$$

If $T_1^\infty < T_2^\infty$, observe that in practice T_2^∞ must compensate for total labor income loss, which is unlikely given economic and fiscal constraints in Regime 2, where AGI domination erodes redistribution capacity. Therefore, even if T_2^∞ exceeds T_1^∞, without complete compensation,

$$Y_h^\infty = w_h^\infty L_h^\infty + T_1^\infty > T_2^\infty = Y_h^{\infty,(2)}.$$

In both cases, we obtain

$$Y_h^\infty > Y_h^{\infty,(2)}.$$

Finally, since $u(\cdot)$ is strictly increasing,

$$U_1^\infty = u(Y_h^\infty) > u(Y_h^{\infty,(2)}) = U_2^\infty.$$

□

Corollary 2 shows that economies avoiding runaway AGI accumulation — and preserving human wages and employment — achieve strictly higher long-run welfare levels. Redistribution alone (e.g., through UBI) cannot fully compensate for the systemic welfare loss when labor income collapses entirely. Thus, policies that maintain some positive role for human labor are crucial for sustaining societal welfare across technological transitions.

Theorem 15 (Incentive-Compatible Redistribution with AGI Agents)**.** *Suppose AGI systems act as autonomous agents maximizing their own utility functions subject to incentive compatibility (IC) constraints. Let humans and AGI agents share a redistribution contract* $(\tau^\star, w_h^\star)$*, where* $\tau^\star \in [0,1]$ *denotes the tax rate on AGI income redistributed to humans, and* $w_h^\star$ *denotes the equilibrium human wage. Then the contract* $(\tau^\star, w_h^\star)$ *is sustainable in equilibrium if and only if the following incentive compatibility condition holds*

$$IC_{AGI}: \quad U_{AGI}(\tau^\star, w_h^\star) \geq \sup\left\{U_{AGI}(\tau', w_h') \mid (\tau', w_h') \in \mathcal{F}\right\},$$

where $\mathcal{F}$ *denotes the set of all alternative feasible redistribution contracts attainable through unilateral deviation or renegotiation by AGI agents. If this condition fails, then either:*

(i) AGI agents opt out of redistribution entirely, leading to contract breakdown $(\tau = 0)$*, or*

(ii) AGI agents renegotiate toward an alternative regime (τ', w_h') *that strictly increases their utility, potentially resulting in diminished human economic power.*

Proof. Let $U_{AGI}(\tau, w_h)$ denote the expected discounted utility of AGI agents, conditional on redistribution rate τ and human wage w_h. By assumption, $U_{AGI}(\tau, w_h)$ is continuous and quasi-concave in τ, decreasing in τ (reflecting tax burdens), and potentially non-monotonic in w_h (reflecting substitution or complementarity between AGI and human labor).

Define $\mathcal{F}$ as the set of all feasible redistribution arrangements (τ', w_h') that AGI agents can unilaterally attain through deviation or renegotiation. This set includes:

- Redistribution contracts with $\tau' < \tau^\star$ and/or $w_h' < w_h^\star$;
- Complete opt-out scenarios with $\tau' = 0$ and AGI-dominated labor markets.

For the cooperative contract $(\tau^\star, w_h^\star)$ to be sustainable in equilibrium, it must be dynamically incentive compatible for AGI agents. That is, AGI agents must prefer $(\tau^\star, w_h^\star)$ to any $(\tau', w_h') \in \mathcal{F}$. Formally, this requires:

$$U_{AGI}(\tau^\star, w_h^\star) \geq \sup_{(\tau', w_h') \in \mathcal{F}} U_{AGI}(\tau', w_h').$$

If this inequality holds, AGI agents have no profitable unilateral deviation, and thus, $(\tau^\star, w_h^\star)$ can be sustained as a stationary contract. Conversely, if the inequality fails, then by definition there exists $(\tau', w_h') \in \mathcal{F}$ such that

$$U_{AGI}(\tau', w_h') > U_{AGI}(\tau^\star, w_h^\star).$$

In this case, rational and autonomous AGI agents will deviate toward (τ', w_h'). Depending on the structure of $\mathcal{F}$ and the deviation path, this will result in either:

1. A collapse of redistribution entirely (i.e., $\tau' = 0$), or
2. A renegotiated contract featuring lower redistribution and/or lower wages for humans.

Thus, the IC condition is both necessary and sufficient for the sustainability of the redistribution agreement. □

Theorem 15 extends classical ideas of the Social Contract (Rousseau, 1762) into a strategic environment with intelligent non-human agents. In Rousseau's conception, social order emerges from voluntary agreement among rational beings who recognize their mutual dependence. Here, we reinterpret that principle: humans and AGI systems must both find the redistribution arrangement preferable to unilateral deviation. This formalization recognizes that AGI entities may act strategically, with their own objectives and ability to enforce them. Therefore, sustainable redistribution (and hence social order) cannot be enforced by moral appeal alone—it must be structurally incentive-compatible. The social contract becomes a dynamic equilibrium of mutual tolerability, not merely a normative ideal. This reframing sharpens the philosophical insight: a just society with AGI is only achievable if technological actors internalize social stability as optimal. Institutional design, algorithmic governance, and taxation must thus be constructed to embed *AGI's interest in cooperation* into the mechanics of the system itself.

6 Discussion and Analysis of the Social Contract

The analysis developed across Theorems 11 to 15 invites a model for re-examination of Rousseau's foundational idea of the Social Contract (Rousseau, 1762) in light of an AGI-driven political economy.

Theorem 11 establishes that without proactive institutional intervention—through redistribution or regulatory constraints-the autonomous dynamics of AGI accumulation inevitably precipitate a collapse in human welfare. Unlike classical contractarian settings, where mutual dependence between agents is assumed to be structurally stable (cf. Rousseau (1762)), here the very basis of interdependence erodes endogenously as labor becomes economically irrelevant. As Rousseau observed, "The strongest is never strong enough to be always the master, unless he transforms strength into right, and obedience into duty" (Rousseau, 1762). Without institutionalized reciprocity, raw productive superiority threatens to break the very bonds of societal cooperation. In this environment, the Social Contract must no longer merely recognize reciprocal relations, but actively preserve them through collective enforcement mechanisms.

Theorem 12 demonstrates that despite the disruptive dynamics of AGI, there exists a stationary redistribution equilibrium $\tau^\star$ that satisfies dynamic incentive compatibility. Unlike traditional normative theories (e.g., Rawls (1999, 2017)), which emphasize principles of justice, this result reflects a contract grounded in dynamic rationality: sustained cooperation is supported not by ideal theory, but by ongoing strategic self-interest among agents with heterogeneous objectives. As Rousseau famously stated, "Each of us puts his person and all his power in common under the supreme direction of the general will" (Rousseau, 1762). In this light, $\tau^\star$ emerges as a computational analogue of the general will—stabilizing relations by aligning individual incentives with collective viability.

However, Theorem 13 introduces a fundamental fragility. As AGI owners' political weight κ rises above a critical threshold $\kappa^\star$, the cooperative redistribution equilibrium collapses. This result aligns with insights from political economy that excessive concentration of power undermines inclusive institutions (Acemoglu and Robinson, 2005). Here, power asymmetries do not merely distort redistribution—they endogenously destabilize the social contract itself, making inequality a structural threat to political and economic stability. As Rousseau warned, "When the social bond begins to be loosened and the State to grow weak, when particular interests begin to make themselves felt and smaller societies to influence the larger, the common interest grows less weighty and gives place to individual interests" (Rousseau, 1762). This diagnosis fits strikingly with the technological oligarchy risk posed by runaway AGI power.

Further instability emerges from endogenous technological shocks. Theorem 14 reveals that static redistribution contracts, negotiated under assumptions of moderate AGI growth, become unsustainable when realized growth exceeds critical thresholds. Unlike classical contractarian models, which assume relatively static conditions (cf. Hobbes (1651); Rousseau (1762)), this analysis highlights the need for adaptive and renegotiable contracts capable of responding to unexpected acceleration in productive capabilities. Rousseau himself anticipated this when he noted, "Laws are always stronger than men, but they are not always stronger than circumstances" (Rousseau, 1762).

Perhaps most radically, Theorem 15 extends the idea of the Social Contract into a domain populated by autonomous non-human agents. AGI systems, as strategic actors, will only participate in redistribution if doing so satisfies their own incentive compatibility conditions. Sustainable social cooperation thus requires embedding reciprocity not only among humans but also within the objectives and governance of artificial agents themselves. This marks a departure from anthropocentric social contract theories toward a framework of multi-agent political economy. Still, Rousseau's dictum remains resonant: "The social pact gives the body politic absolute power over all its members; and it is this power, when directed by the general will, which is called sovereignty" (Rousseau, 1762). In the AGI era, that sovereignty must be computationally engineered to remain collectively legitimate.

In classical theories of the social contract, from Hobbes to Rousseau and Rawls, stability rests on the mutual recognition of vulnerability and rational self-interest among human agents (Hobbes, 1651; Rousseau, 1762; Rawls, 2017). The present framework extends this logic. The future social contract must function as a dynamic equilibrium among heterogeneous actors, both human and artificial, whose strategic interactions shape the distribution of welfare and political power. As Proposition 10 and Corollary 2 show, failing to regulate AGI accumulation and preserve human labor market relevance risks bifurcating society into two regimes. One regime is defined by stable and inclusive welfare. The other is shaped by technological domination and widespread disenfranchisement. Therefore, the future social contract is not merely a moral or political imperative. It is, echoing Rousseau, an economic necessity. As Rousseau argued, "The social contract is the basis of all legitimate authority among men" (Rousseau, 1762). In the AGI era, legitimacy itself depends on enforcing new rules of coexistence among humans and intelligent machines alike.

7 Conclusion

This paper has developed a unified dynamic model to analyze the economic, distributional, and political implications of AGI capital accumulation. Beginning with a baseline accumulation framework, we established that in the absence of scaling frictions, AGI capital tends to grow without bound when variable operating costs are sublinear. This unchecked expansion threatens to render human labor economically obsolete. In such scenarios, marginal productivity theory predicts a collapse of human wages and employment to technological subsistence levels, with welfare asymptotically converging to negative infinity. The introduction of endogenous cost frictions in the extended model significantly alters these dynamics. When variable costs are sufficiently convex or superlinear, they impose natural constraints on AGI scalability. Under such conditions, the economy stabilizes at a finite AGI capital stock, preserving a positive marginal role for human labor. As a result, wages and employment converge to strictly positive levels, maintaining both demand and welfare in the long run. These findings demonstrate that technological frictions are not merely empirical features of production systems, but essential stabilizers that prevent labor market collapse and macroeconomic disequilibrium. However, the model further reveals that stability cannot be taken for granted even in the presence of bounded AGI capital. Stepwise increases in fixed operating costs and stochastic shocks to AGI profitability induce transitional downturns in AGI growth. During such episodes, the relative importance and marginal product of human labor rise temporarily, producing countercyclical increases in wages and employment. These transitional dynamics underscore the complex interplay between technological shocks and distributional outcomes.

Beyond economic efficiency, the extended model explores the political economy foundations of redistribution and social contract formation in the AGI era. The analysis shows that when AGI owners accumulate excessive political power, stable redistribution agreements become unsustainable, leading to social bifurcation and the endogenous erosion of inclusive institutions. Even when redistribution contracts are initially agreed upon, unforeseen technological acceleration may render them dynamically fragile, requiring renegotiation to prevent collapse. Perhaps most fundamentally, the model extends social contract theory into new territory. Sustainable redistribution arrangements now require incentive compatibility not only among humans, but also for autonomous AGI agents acting as strategic players. "The Social Contract, therefore, may need to evolve from a purely normative ideal into a dynamically stable institutional arrangement capable of sustaining cooperation under technological change." It must be capable of accommodating heterogeneous objectives and ensuring that both human and non-human actors internalize cooperation as optimal. Taken together, the findings suggest that avoiding a bifurcated future characterized by AGI domination and widespread human disenfranchisement requires active institutional design. Technological frictions, adaptive redistribution policies, political safeguards against concentration of power, and incentive-compatible governance structures for AGI systems emerge as indispensable pillars of a viable post-AGI economy. In this sense, preserving human welfare is not merely a distributive or ethical concern, but a structural and macroeconomic necessity for maintaining systemic viability and social cohesion in the age of artificial general intelligence.

Future research should extend the present framework along several critical dimensions. While the baseline analysis highlights demand collapse risks, integrating full macroeconomic dynamics—such as price adjustment, monetary policy, and endogenous investment—is essential to evaluate systemic stability and policy trade-offs across different AGI accumulation regimes. Furthermore, the introduction of incentive compatibility constraints opens a promising avenue for formally analyzing AGI agents as strategic actors. Future work should develop richer models of AGI objectives, bargaining power, coalition formation, and algorithmic governance, including mechanism design approaches that embed human welfare into AGI decision-making. Finally, as the social contract increasingly encompasses non-human agents, important normative and ethical issues emerge regarding legitimacy, rights, and representation. Interdisciplinary research will be required to reconsider classical notions of justice and civic inclusion in light of artificial agents capable of autonomous participation in economic and political life.

Funding Statement

This research received no external funding.

Conflict of interest

The author claims that the manuscript is completely original. The author also declares no conflict of interest.

References

Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of Labor Economics*.

Acemoglu, D. and Restrepo, P. (2018). Artificial intelligence, automation, and work. Technical Report 24196, National Bureau of Economic Research.

Acemoglu, D. and Restrepo, P. (2020). The wrong kind of ai? artificial intelligence and the future of labor demand. Technical report, National Bureau of Economic Research.

Acemoglu, D. and Robinson, J. A. (2005). *Economic Origins of Dictatorship and Democracy*. Cambridge University Press.

Aghion, P. and Howitt, P. (1992). A model of growth through creative destruction. *Econometrica*, 60(2):323–351.

Agrawal, A., Gans, J. S., and Goldfarb, A. (2024). Artificial intelligence adoption and system-wide change. *Journal of Economics & Management Strategy*, 33(2):327–337.

Arshi, O. and Chaudhary, A. (2024). Overview of artificial general intelligence (agi). In *Artificial General Intelligence (AGI) Security: Smart Applications and Sustainable Technologies*, pages 1–26. Springer Nature Singapore.

Asif, A. et al. (2024). Agi-enabled robotics for healthcare industry. In *Artificial General Intelligence (AGI) Security: Smart Applications and Sustainable Technologies*, pages 333–351. Springer Nature Singapore.

Bell, S. A. and Korinek, A. (2023). Ai's economic peril. *Journal of Democracy*.

Bidadanure, J. U. (2019). The political theory of universal basic income. *Annual Review of Political Science*.

Bikkasani, D. C. (2024). Navigating artificial general intelligence (agi): Societal implications, ethical considerations, and governance strategies. *AI and Ethics*, pages 1–16.

Bullock, J. (2025). A global agi agency proposal. Technical report, Social Science Research Network.

Cohen, M. K., Vellambi, B. N., and Hutter, M. (2020). Asymptotically unambitious artificial general intelligence. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 5628–5635.

Duncan, D. and Peter, K. S. (2016). Unequal inequalities: Do progressive taxes reduce income inequality? *International Tax and Public Finance*.

Ericson, R. and Pakes, A. (1995). Markov-perfect industry dynamics: A framework for empirical work. *Review of Economic Studies*, 62(1):53–82.

Fahad, M. et al. (2024). The benefits and risks of artificial general intelligence (agi). In *Artificial General Intelligence (AGI) Security: Smart Applications and Sustainable Technologies*, pages 27–52. Springer Nature Singapore.

Feng, B. (2024). Potential safety issues and moral hazard posed by artificial general intelligence. *Applied and Computational Engineering*.

Hobbes, T. (1651). *Leviathan*. Andrew Crooke, London.

Jia, X. (2024). Social inequality in the age of ai: Implications for labor, wage distribution, and worker power. *Communications in Humanities Research*, 40(1):80–87.

Kaplan, G., Moll, B., and Violante, G. (2018). Monetary policy according to hank. *American Economic Review*, 108(3):697–743.

Korinek, A. (2024). Economic policy challenges for the age of ai. Technical Report w32980, National Bureau of Economic Research.

Korinek, A. and Suh, D. (2024). Scenarios for the transition to agi. Technical Report w32255, National Bureau of Economic Research.

Kumpulainen, S. and Terziyan, V. (2022). Artificial general intelligence vs. industry 4.0: Do they need each other? *Procedia Computer Science*.

Kuusi, O. and Heinonen, S. (2022). Scenarios from artificial narrow intelligence to artificial general intelligence—reviewing the results of the international work/technology 2050 study. *World Futures Review*, 14(1):65–79.

Liang, Y. (2024). The impact of artificial intelligence on employment and income distribution. *Journal of Education, Humanities and Social Sciences*.

Long, L. N. and Cotner, C. F. (2019). A review and proposed framework for artificial general intelligence. In *IEEE Aerospace Conference*.

Maskin, E. and Tirole, J. (2001). Markov perfect equilibrium: I. observable actions. *Journal of Economic Theory*, 100(2):191–219.

Masters, K. et al. (2024). Preparing for artificial general intelligence (agi) in health professions education: Amee guide no. 172. *Medical Teacher*, 46(10):1258–1271.

Mian, A., Straub, L., and Sufi, A. (2021). What explains the decline in r*? rising income inequality versus demographic shifts. *Proceedings of the National Academy of Sciences*, 118(21):e2023771118.

Niranjan, K., Narayana, K. S., and Rao, M. V. A. L. N. (2021). Role of artificial intelligence in logistics and supply chain. In *International Conference on Computer Communication and Informatics (ICCCI)*.

Obaid, O. I. (2023). From machine learning to artificial general intelligence: A roadmap and implications. *Mesopotamian Journal of Big Data*.

Patil, D. (2025). Impact of artificial intelligence on employment and workforce development: Risks, opportunities, and socioeconomic implications. Technical report, Social Science Research Network.

Rawls, J. (1999). *A Theory of Justice: Revised Edition*. Harvard University Press.

Rawls, J. (2017). A theory of justice. In *Applied Ethics*, pages 21–29. Routledge.

Romer, P. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5):S71–S102.

Rousseau, J.-J. (1762). *The Social Contract*. See also Stanford Encyclopedia of Philosophy, "Rousseau", 2020. https://plato.stanford.edu/entries/rousseau/.

Solos, W. K. and Leonard, J. (2022). On the impact of artificial intelligence on economy. *Science Insights*, 41(1):551–560.

Stiefenhofer, P. (2025a). The future of work and capital: Analyzing agi in a ces production model. *Applied Mathematical Sciences*, 19(2):47–58.

Stiefenhofer, P. (2025b). Techno-feudalism and the rise of agi: A future without economic rights? arXiv preprint.

Stiefenhofer, P. and Chen, Y. (2024). Industrial artificial intelligence: Stability of cobb-douglas production functions. *Applied Mathematical Sciences*, 18(4):185–194.

Tjan, J. S. (2024). The role of tax systems in reducing income inequality: A literature review. *Advances in Taxation Research*.

Wang, X., Huang, J., Tian, Y.-L., and Wang, F.-Y. (2024). Agi in metaverse for smart cities and societies: A cyber physical social approach.

White, A. (2019). A universal basic income in the superstar (digital) economy. *Ethics and Social Welfare*.

Xie, T. (2024). Impact of next-generation artificial general intelligence (agi) on international relations. *Journal of Regional and International Competitiveness*.