We Wont be Missed:
Work and Growth in the Era of AGI

Pascual Restrepo
pascual.restrepo@yale.edu
Yale University

July 4, 2025

Abstract

This chapter explores theoretically the long-run implications of Artificial General Intelligence (AGI) for economic growth and labor markets. AGI makes it feasible to perform all economically valuable work using compute. I distinguish between bottleneck and accessory work—tasks essential vs. non-essential for unhindered growth. As computational resources expand: (i) the economy automates all bottleneck work, (ii) some accessory work may be left untouched by AI and assigned exclusively to humans, (iii) output becomes linear in compute and labor and its growth is driven by the expansion of compute, (iv) wages converge to the opportunity cost of computational resources required to reproduce human work, and (v) the share of labor income in GDP converges to zero.

This chapter studies the long-run behavior of wages and growth in an economy where *Artificial General Intelligence* (AGI) is developed and computational resources increase over time. AGI allows the economy to complete all relevant work using computing systems. These systems consume computational resources but do not require human input, guidance, or effort to accomplish work.

The key economic problem is how to allocate finite (but growing) computational resources and human labor to accomplishing the work needed to produce output. The chapter introduces a key distinction between bottleneck and accessory work.

- *Bottleneck work* comprises tasks essential for economic growth. Output cannot expand indefinitely unless inputs in bottleneck tasks also expand or become infinitely valuable.
- *Accessory work* is non-essential to growth. Output can expand indefinitely even if these tasks are discarded or limited in input.

My main theoretical result shows that *all* bottleneck work is eventually automated while some accessory work may be left untouched by AI. Once this occurs, output shifts from being multiplicative in compute and human effort to being additive, and the long-run growth rate of the economy is pinned by the growth rate of compute.

Despite AI completing all bottleneck tasks without human input, people still hold jobs. They can contribute by fulfilling bottleneck work. Human labor remains valuable because it saves scarce computational resources. Alternatively, workers may perform accessory work, where it is impractical to use compute since we already have too many workers. In the first case, wages are pinned by the value of computational resources saved. In the later, wages are bounded above by the value of computational resources it would cost to automate accessory work.

In sum, the advent of AGI changes the way labor is valued. Before AGI, wages reflected the importance of bottleneck work and the scarcity of labor with the requisite skill for this work. With AGI, wages reflect the computational cost of replicating the work produced by all human labor. Despite the fact that human labor retains some value, its contribution to GDP and growth becomes vanishingly small, with the share of labor in GDP converging to zero, and all income being eventually accruing to compute.

I then expand the analysis to an economy where AGI can be used to complete scientific work, accelerating the pace of technological progress. Without AGI for science, technological progress is constrained by population growth (as in the semi-endogenous growth models of Jones, 1995;

Kortum, 1997; Segerstrom, 1998). With AGI, all scientific bottleneck work is automated and the rate of technological progress is determined by the growth rate of compute. This may generate sustained exponential growth despite shrinking population, but does not create a singularity or infinite growth explosion.

The analysis here complements existing work on the economics of AI, including several contributions in a previous NBER volume (see Aghion, Jones and Jones, 2019; Acemoglu and Restrepo, 2019). Methodologically, I build on ideas from task models, which center on the problem of how to accomplish multiple work tasks using different production techniques (Autor, Levy and Murnane, 2003; Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018, see). As in these models, the core economic problem I study is that of allocating work to human labor or AGI systems, and deriving the implications of this process for wages and output. The focus on the transition to AGI is shared with complementary work by Anton Korinek and collaborators that influenced some of my thinking (see Trammell and Korinek, 2023; Korinek and Suh, 2024). Outside of traditional economics, there is a large and interesting literature on world building exercises imagining how the economy and world are transformed by AGI (see Kokotajlo et al., 2025; Drago and Laine, 2025).

The chapter is organized as follows. Section 1 presents the analysis in a production economy where AGI can be used to produce all production work with compute. Section 2 then extends the analysis to a semi-endogenous growth economy where AGI can be used for scientific work. Section 3 concludes with a discussion of policy and issues related to the distribution of income and meaningful work.

1 Production and Work with Artificial General Intelligence

The economy produces output Y_t by completing *work*—the use of human or computational resources to accomplish tasks needed to generate valuable output. The set of all valuable work is Ω, with each specific type of work indexed by ω. I assume Ω is a finite (but large set).[1]

The quantity of work ω completed at time t is given by

$$X_t(\omega) = L_t(\omega) + \frac{1}{\alpha_t(\omega)} Q_t(\omega). \tag{1}$$

[1]Infinitesimal similar results but technical complications

Work ω can be accomplished by human labor $L_t(\omega)$ or using computational resources $Q_t(\omega)$. The quantity of work produced by a human per unit of time is normalized to 1. Emulating or replicating this work using computers costs $\alpha_t(\omega)$ units of compute. We could have $\alpha_t(\omega) = \infty$ so that it is impossible to replicate human effort in ω with computers.

The quantity of output produced is then

$$Y_t = F(\{X_t(\omega)\}_{\omega \in \Omega}),$$

where F is increasing in all entries, differentiable, concave, and exhibits constant returns to scale. I also assume marginal products converge in $[0, \infty]$ for any converging sequence of inputs in $[0, \infty]$.

There are different human skills, indexed by $s \in \mathcal{S}$, and in quantity $H(s)$. People of skill s can accomplish work in $\Omega(s) \subseteq \Omega$, where the $\Omega(s)$ partition Ω. This implies

$$\sum_{\omega \in \Omega(s)} L_t(\omega) \leq H(s).$$

The partition $\Omega(s)$ captures the degree to which skills can only be used for a given type of work ($\Omega(s)$ is a one-to-one mapping from $\mathcal{S}$ to Ω) or they are general ($\Omega(s) = \Omega$ and there is a single universal skill). The assumption that there is no overlap between $\Omega(s)$'s simplifies exposition.

The total quantity of computational resources at our disposal is Q_t. Think of this as the total number of computations the economy can perform per unit of time, given its data centers and chips. The *computational resource constraint* of the economy is then

$$\sum_{\omega \in \Omega} Q_t(\omega) \leq Q_t.$$

The Premise: The objective here is to characterize the behavior of an economy facing two forms of technological progress: the development of *Artificial General Intelligence* and the increased expansion of its computing capabilities. What happens in an economy that reaches AGI and acquires enough compute to run it?

The first form of technological progress considered is the development of AGI. This premise can be expressed in terms of the work equation (1) as:

Premise 1 (The economy develops AGI.)**.** *For all work $\omega \in \Omega$, $\alpha_t(\omega)$ reaches a finite value by date $T(\omega)$*

and converges to a terminal value $\alpha(\omega) \in (0, \infty)$ from there on.

AGI is the knowledge or technology for transforming raw compute into all types of useful work. Acquiring AGI means we figured out how to train or create computer systems capable of performing all work currently accomplished by humans. This definition distinguishes between AGI and computational resources. AGI is the recipe of how to use compute—the input—to accomplish any type of work—the output. Doing so consumes some of our finite computational resources, Q_t.[2]

While I take this premise as a point of departure in my analysis, it is helpful to clarify three points about what is being assumed here.

- **Feasible vs Realized:** The definition of AGI states that, in principle, we can replicate what people do with enough compute. Depending on the computing costs $\alpha(\omega)$ of replicating what a person does we may decide not to use computational resources to accomplish this work. The key point is that AGI makes it feasible to do so; just because we develop AGI does not mean it will be practical to carry out all work with computers, since our computational resources are finite and can be applied to more profitable uses.

- **Physical work:** Some work requires interacting physically with the world. The premise here is that, when needed, computer systems can control machine and hardware to accomplish this work. When appropriate, I re-interpret Q_t as a bundle of energy and computational resources that computers use to carry out both cognitive and physical work.

- **Social work:** one may argue some work requires social interaction and must be carried out by humans. The "human touch" and "empathy" of a therapist or healthcare provider may be impossible to replicate, creating a premium for work completed by people. The assumption behind Premise 1 is that *quantity has a quality all its own*: any deficit of computer vs humans may be overcome by the sheer application of compute to the task at hand. Imagine, for example, an AI system that perfectly emulates the best therapists in the world, is at your disposal at any time, knows you perfectly, never gets tired, and is open to use 100% of its intellect and capabilities to help you overcome your anxieties. The system may be too costly to be practical,

[2]The definition used here differs from other concepts, such as *Transformative AI* (defined as "AI developments that significantly transform the economy") or *Super-intelligence* (AGI that keeps recursively self-improves itself, reaching levels of intelligence that far exceed human ones). The definition of AGI used here does not need AI to be super-intelligent.

but would you insist on visiting a human therapist when you could instead be treated by this superhuman team of world experts? Imagine an AI system that uses vast amounts of compute to diagnose and treat medical conditions. Would you keep a sick children from accessing this medical attention because you insist AI lacks "the human touch"? If valid, this argument implies that we can still represent such situations by high values of $\alpha(\omega)$, which capture the extra computing resources needed to compensate for the "human touch". Whether it is practical or not to perform such work with compute is a different question addressed below.

The second form of technological progress involves increased computational resources, Q_t. The premise here is that computational resources are finite at each point in time but grow over time and becomes arbitrarily abundant.

Premise 2 (Abundant Compute.)**.** *Computational resources Q_t are finite but grow without bound in time.*

This premise is motivated by historical trends and regularities, such as Moore's law. The premise is that we will continue to expand computational resources in the future, in the same way we have expanded them historically, with no obvious ceiling in the near or mid term and a limit that far exceeds current computational resources.

Historically, we have greatly expanded computing capabilities since the creation of modern computers and transistors in 1970. Computational resources can be measured in *flops*—the total number of floating-point operations per second that all computers in the economy could collectively perform. From 1980 to 2007, we increased compute by three orders of magnitude, from 10^{15} to 10^{18} *flops*. Currently, our economy has a peak capacity of 10^{21} *flops*. It is estimated that this could increase to 10^{54} *flops* in the long run, suggesting that there is plenty of room for compute to grow. For reference, a human brain is estimated to perform 10^{16}–10^{18} *flops*, so computational resources are already en route to exceed human brainpower by orders of magnitude.[3]

Progress in computing capabilities is assumed exogenous and is subsumed in the path for Q_t. As a benchmark, one could imagine computational resources Q_t growing exponentially in time—as in some variants of Moore's law—, though my results only require Q_t to become sufficiently large.

[3]See Hilbert and López (2011) for trends in compute over 1980–2007. See AI Impacts (2023) for estimates of current compute. See Bostrom (2003) and Bostrom (2014) for estimates of future computational resources. See Sandberg and Bostrom (2008) and Open Philantropy (2020) for estimates of the computing power of the human brain.

Bottlenecks: In what follows, I let $F(\{X(\omega)\})$ denote the output obtained when the quantity of work is $X(\omega)$ and $F_\omega(\{X(\omega)\}) > 0$ denote the marginal gain of work ω.

A core component of the analysis is defining bottleneck and accessory work.

Definition 1. *Work ω is bottleneck if, for any $\{X_t(\omega)\}$ such that $F(\{X_t(\omega)\})$ is unbounded, either i. $X_t(\omega)$ is unbounded or ii. $F_\omega(\{X_t(\omega)\})$ is unbounded.*

The definition captures an intuitive notion of a bottleneck: a type of work that is necessary for sustained growth, in the sense that it must expand or its price would inflate to exorbitant levels, capturing an expanding share of output. Potential examples of bottleneck work includes feeding and sheltering people, producing energy, maintaining the productive infrastructure of the economy, advancing science, decision making, logistics and delivery, and maintaining national security as well as policing nations to preserve order and stability. These types of work are mission critical in the sense that is hard to imagine an economy that keeps growing in a sustained way but lacks any of these components.

Definition 2. *Work ω is accessory if, there is $X_t(\omega)$ such that $F(\{X_t(\omega)\})$ is unbounded while both $X_t(\omega)$ and $F_\omega(\{X_t(\omega)\})$ are bounded.*

The definition shows that accessory work is the opposite of a bottleneck: we can grow the economy while keeping this type of work fixed and without commanding too high of a price. Potential examples include work associated with arts and crafts, performing for others, literature, and hospitality and fine dining. We may also have design and customer support work, which may not scale with the economy either. Finally, we may have judicial work, as well as work associated with religious and civic organizations, as well as nature preservation groups. Even the work performed by academic economists may prove to be accessory, as it is unlikely to become ever more valuable in the future. These are just possible examples, as the exact partition of what turns out to be a bottleneck or accessory work depends on future preferences, the structure of production, and the key existential problems and challenges faced by future people.

1.1 Limit Behavior

This subsection characterizes the limit behavior of the economy as $t \to \infty$ under Premises 1 and 2. I characterize the properties of a competitive equilibrium: an allocation of compute and human labor that maximizes output and where factors of production are paid their marginal products. This choice highlights the key economic forces at play.

Proposition 1. *All bottlenecks are eventually automated while some accessory work may be left to labor.*

The proposition clarifies how work is carried out in the AGI economy. All bottleneck work is eventually automated and produced with compute, while some accessory work is left exclusively to humans. Human labor may still produce some bottleneck jointly with AGI, adding to the quantity of work obtained, or may specialize fully in accessory work. In the special case where all work is a bottleneck, the proposition implies that *all* work is eventually automated and produced with compute.

This last result seems to counter Ricardo's principle of *Comparative Advantage*. We can think of the AGI economy as a world where we engage in trade with "a country of geniuses in a data center," exchanging compute for productive work.[4] Wouldn't Ricardo's principle imply that the AGI and human countries should fully specialize to maximize the gains from trade? Why is it optimal for AGI to produce all bottleneck work?

The reason why this logic breaks is instructive and provides an heuristic proof of the proposition. Suppose AGI specializes in the production of a subset of work with compute while human labor specializes in the remaining components. Over time, the economy will become unbalanced, producing an expanding quantity of the first type of work and a fixed quantity of the later. This imbalance cannot be optimal: to keep the economy growing one must expand *all* bottleneck work at the same rate. The only circumstance in which such unbalanced path can be optimal is one where the non-automated work is accessory, as claimed in the proposition.[5]

Let's now characterize the behavior of output and wages once we reach the point when all bottleneck work is automated. Define the compute-equivalent units (CEU) of skill s as

$$\text{CEU}(s) = \max_{\omega \in \Omega(s)} \alpha(\omega).$$

This gives the computational resources needed to replicate work carried by people of skill s in the

[4]This apt metaphore is from Anthropic's Amodei (2024).

[5]The idea that Ricardo's principle of Comparative Advantage calls for *full* specialization is a common misunderstanding. Ricardo's original 2×2 example predicts that *at least one* of the two countries specializes (see chapter 2 in Feenstra and Taylor, 2017, for a textbook treatment). When one country is large and endowed with enough resources (the AGI country in our economy), the equilibrium involves the large country producing both goods and determining their relative prices. The small country (humans), on the other hand, specializes in a single good, as in Proposition 1.

most computationally complex task they perform. Premise 1 implies CEU(s) is finite for all s.

Proposition 2. *Output converges to*

$$Y_t \;=\; A\,\big(Q_t \;+\; \sum_{s\in\mathcal{A}} CEU(s)\,H(s) \;+\; \sum_{s\in\mathcal{N}} \widetilde{CEU}(s)\,H(s)\big), \tag{2}$$

for some $\widetilde{CEU}(s) < CEU(s)$*, where* $\mathcal{A}$ *is the set of skills performing only automated work and* $\mathcal{N}$ *the set of skills that perform accessory work not automated.*

Aggregate production possibilities in the economy become additive in compute and labor, with labor expressed in units of compute. This holds for *any* initial production function F. For example, if the different forms of work are combined a-la Cobb-Douglas, we would go from an economy where compute and human labor are combined in a multiplicative way (before AGI is developed) to one where they are combined in an additive way (once AGI is developed and we acquire enough compute).

To understand the result, let's consider the special case where all work is a bottleneck. Proposition 1 implies all work is automated, which means compute is used for all existing work. At this point, we can write output (given an allocation of labor) as

$$Y_t \;=\; \max_{Q_t(\omega)} \; F\big(\{L_t(\omega) \;+\; \frac{1}{\alpha(\omega)}\,Q_t(\omega)\}\big) \quad \text{s.t:} \sum_{\omega\in\Omega} Q_t(\omega) \leq Q_t.$$

Let $\widetilde{Q}_t(\omega) \;\equiv\; Q_t + \alpha_t(\omega)\,L_t(\omega)$ denote effective resources (in units of compute) allocated to work ω. We can rewrite the maximization problem as

$$Y_t \;=\; \max_{\widetilde{Q}_t(\omega)} \; F\big(\{\frac{1}{\alpha_t(\omega)}\,\widetilde{Q}_t(\omega)\}\big) \quad \text{s.t:} \sum_{\omega\in\Omega} \widetilde{Q}_t(\omega) \leq Q_t + \sum_{\omega\in\Omega} \alpha_t(\omega)\,L_t(\omega),$$

where the economy maximizes output subject to a *total resource constraint* pooling compute and human labor. Constant returns to scale implies a solution of the form

$$Y_t \;=\; A\,\big(Q_t + \sum_{\omega\in\Omega} \alpha_t(\omega)\,L_t(\omega)\big),$$

for some $A > 0$ equal to the rate of transformation of compute into output. To conclude, let's turn to labor. The allocation of labor that maximizes total resources assigns all labor of skill s to $\arg\max_{\omega\in\Omega} \alpha(\omega)$, which yields

$$Y_t \; = \; A \, \big(Q_t + \sum_{s\in\mathcal{S}} \mathrm{CEU}(s)\, H(s)\big).$$

These steps clarify the economic logic behind Proposition 2. AGI allows us to produce all bottleneck work with compute. At this point, compute pins the value of work: having a worker producing one unit of work ω is the same as having $\alpha(\omega)$ extra units of compute, which is the same as having $A\ \alpha(\omega)$ units of output. The best thing labor can do is to specialize in the work that saves the most compute, which frees computational resources capable of generating an output A CEU(s).

The above reasoning must be modified slightly by the existence of accessory work. Accessory work that is automated can be handled as above. If some accessory work is not automated, it is because it is not worth wasting scarce compute on it. Suppose ω is performed by workers of skill $s \in \mathcal{N}$ and not worth automating. Reallocating $\alpha(\omega)$ units of compute from producing bottlenecks (valued at $A\ \alpha(\omega)$) to producing ω (valued at MPL(s)—the marginal product of skill s labor) must lower output, which implies

$$A\ \alpha(\omega) > \mathrm{MPL}(s) \quad \Rightarrow \quad A\ \mathrm{CEU}(s) > \mathrm{MPL}(s).$$

This inequality shows that the value of accessory work performed by skill s labor is bounded by A CEU(s) and must then eventually converge to an additive constant (this last claim is a nontrivial implication of constant-returns to scale in F).[6]

Proposition 2 shows how the automation of bottlenecks allows output to scale with compute, sustaining economic growth.

Proposition 3. *Output grows at the same rate as computing resources Q_t.*

This provides cause for optimism: In a world of AGI, the economy can grow simply by expanding computational capabilities. No other form of technological progress is needed after AGI makes it feasible to turn compute into all kinds of work. This is true even if there is accessory work left unautomated, as this work does not hinder growth.

[6] A subtle detail behind the statement of Proposition 2 is that human workers fall in two groups. Workers of skill s for whom one element of $\omega^*(s)$ is automated perform work only in $\omega^*(s)$ and all such work is automated. This is group $\mathcal{A}$. The remaining workers are group $\mathcal{N}$.

How is the growing income distributed? In a competitive economy, workers and compute are paid their marginal contribution to output. These can be read right away from the expression for output in Proposition 2.

Proposition 4. *The real price of computing resources converges to A and real wages to*

$$W(s) = \begin{cases} A \; CEU(s) & \text{if } s \in \mathcal{A} \\ \\ A \; \widetilde{CEU}(s) & \text{if } s \in \mathcal{N} \end{cases}$$

In an AGI economy, workers are paid the value of the compute needed to replicate their work. Human skill remains valuable—even in a world where all work is automated—because it can be used to accomplish useful work, saving scarce computational resources. This hints at an important distinction: AGI does not render labor redundant; it makes it replicable through computation. Since compute is scarce, human labor retains value, given by the opportunity cost of deploying computational resources.

Skills that perform accessory work command a discounted wage, in the sense that $\widetilde{\text{CEU}}(s) < \text{CEU}(s)$. Accessory work left to humans is not a growing source of riches. The reason why this work is left untouched is that we already have too many workers to do it. This lowers its value (hence the discount) and makes it impractical to automate it.

The significant result is that wages become fully decoupled from growth, which in the AGI economy is entirely driven by expanding computing resources. Despite the fact that the economy keeps expanding over time, the value of human labor stops growing and remains bounded above by the value of compute needed to replicate it. Because wages remain bounded while the economy continues to expand, all income eventually accrues to compute.

Proposition 5. *The share of compute in GDP converges to 1 and the share of labor in GDP to zero.*

The share of labor in GDP can be bounded by the share of human compute in our total computational resources, measured by

$$\text{Share human compute}_t \equiv \frac{\sum_{s \in \mathcal{S}} \text{CEU}(s) \; H(s)}{Q_t + \sum_{s \in \mathcal{S}} \text{CEU}(s) \; H(s)}.$$

This converges to zero over time. The numbers on computational resources cited above give a sense of magnitudes. The value of human compute, $\sum_s \text{CEU}(s)\, H(s)$, is measured in billions (population) times $10^{15} - 10^{18}$ *flops*. The value of total compute in the economy, on the other hand, could be as high as 10^{54} *flops*, making human compute meager in comparison. These calculations show that in an AGI economy where labor paid its compute-equivalent value, the labor share will plummet, as total compute is projected to far exceed human compute. Conversely, most if not essentially *all* income will accrue to owners of computing resources.

The fact that the value of human labor remains capped and shrinks as a share of output does not imply that AGI made society and workers poor in any sense. Adding up the income generated by compute and human labor, society becomes richer (and keeps getting richer over time so long as it expands computing resources!). Even if we leave aside the income accruing to compute, the transition from the pre-AGI to the AGI economy necessarily makes labor income more valuable as a whole.

Proposition 6. *The sum of all people wages is higher in the limit with AGI than in the pre-AGI economy (with $Q_t = 0$ and $\alpha_t(\omega) = \infty$).*

The proposition shows that workers *as a whole* benefit from transitioning to the AGI economy. Why? The arrival of AGI cannot make us worst off collectively, since we could always set up a no-AGI zone and continue carrying our lives as if nothing had happened. This delivers at least the same wages we had before.

The exact argument is as follows: competitive markets imply that the economy arranges production efficiently under AGI. There is no rearrangement that can raise output. Suppose we take 1% of human workers and send them to a non-AGI zone. These workers produce $(W_{\text{pre-AGI}} \cdot H) \times 1\%$ in the non-AGI zone (an implication of constant-returns to scale), but we give up their wages $(W_{\text{post-AGI}} \cdot H) \times 1\%$ in the post-AGI world. Because this rearrangement cannot raise output, the cost $(W_{\text{post-AGI}} \cdot H) \times 1\%$ exceeds the gains $(W_{\text{pre-AGI}} \cdot H) \times 1\%$ and wages in the post-AGI world are higher than those in the pre-AGI one.

There are limits to the above argument. If the AGI economy uses some finite resource that is also needed by human labor or there are diminishing returns to scale, the above argument breaks down. Still, the above logic shows that the scope for AGI to make human workers collectively less valuable is limited. The "problem" with AGI is that human labor seizes to be a source of *growing* riches, not that it looses value.

1.2 Transition to AGI [Preliminary]

In general, the transition from the current economy to the AGI limit is hard to trace and depends on how fast compute advances (Q_t) relative to progress in AGI (the α's).

To simplify the exposition, assume all work is a bottleneck and is eventually automated. Assume also that $\alpha_t(\omega)$ jumps from infinity to its limit value $\alpha(\omega)$ at some finite time $T(\omega)$.

Two polar cases stand out. On one extreme we have a transition where technology binds and work ω is automated at $T(\omega)$, the exact moment it becomes feasible to do so. This means compute is abundant and ready to be used in any bottleneck tasks that can be automated. On the other extreme, we have a transition where compute binds and work ω is only automated at a future date $\tilde{T}(\omega)$, once the economy acquires enough compute to justify this application.

Proposition 7. *Suppose compute binds along the transition. The marginal value of compute A_t decreases and converges monotonically to $A > 0$. Moreover, for every skill s:*

1. *The work in $\Omega(s)$ is automated sequentially, from least to computational to highest computational requirements.*

2. *Wages $W_t(s)$ decreases gradually around $\tilde{T}(\omega)$ at the same rate as A_t for an interval of time and then continue growing until the next date is reached.*

3. *At the time work in $\arg\max_{\omega\in\Omega} \alpha(\omega)$ is finally automated, wages $W_t(s)$ must equal A_t CEU(s) and decrease gradually from there on at the same rate as A_t.*

As a whole, total wages weakly increase during the transition.

When compute binds, the transition to the AGI limit is smooth and gradual. For every skill s, wages are in general growing over time. This pattern is punctuated by short-lived episodes around the time some of the work performed is automated. During these events, workers gradually reallocate away from the automated work, which is eventually completely taken over by AGI. This process continues until only work in $\arg\max_{\omega\in\Omega} \alpha(\omega)$ remains. Once this work is finally automated, workers are stuck here and their wages are pinned down by the value of compute they save. From there on, their wages decrease at the same rate as A_t until they converge to the values in Proposition 4. On net, this process generates a growing path for total wages, and no sharp declines or "jumps" in the wage of workers.

Proposition 8. *Suppose technology binds along the transition. The marginal value of compute A_t may be non-monotonic and jump up at points during the transition before converging to A. Moreover, for every skill s:*

1. *The work in $\Omega(s)$ is automated at $T(\omega)$.*
2. *Wages $W_t(s)$ may jump down at $T(\omega)$ for some of the $\omega \in \Omega(s)$ if displacement effects dominate.*
3. *Wages $W_t(s)$ necessarily jump down when the last work in $\Omega(s)$ is automated*
4. *Wages may exceed $A_t \cdot CEU(s)$ at various points in the transition.*

As a whole, total wages may jump down at points during the transition and will otherwise increase.

When technology binds, the transition to the AGI limit is jagged, uncertain, and haphazard. The proposition describes a world where compute is abundant and ready to be used the second an application arrives. Technology firms are experimenting in the back with applications of AI. The moment an application pans out, the labor market is disrupted, causing the immediate displacement of some of the people working at the automated tasks. Those who happen to be employed in applications that prove harder to automate, on the other hand, will see rising wages, far exceeding what they will earn in the AGI limit. The flip-side is the looming realization that at some future date, this high wages will evaporate, as this is a premium for the random fact that the work they do just happened to be the last experiment to pan out.

2 Scientific Work

Above I considered an economy where computers and people carry out the work needed for production. Let's now expand the analysis to account for scientific work. This type of work does not deliver output, but expands our knowledge on how to produce more efficiently in the future.

The quantity of output produced is now

$$Y_t = Z_t \, F(\{X_t(\omega)\}_{\omega \in \Omega}),$$

where Z_t is the level of technological sophistication of the economy, improved by completing scientific work. As in semi-endogenous growth models, this evolves according to

$$\frac{\dot{Z}_t}{Z_t} \;=\; Z_t^{-\beta}\, G(\{X_t(\sigma)\}_{\sigma\in\Xi}),$$

where σ denotes scientific work and Ξ is the set of all such work. The function G is increasing in all entries, differentiable, concave, and exhibits constant returns to scale. The elasticity $\beta > 0$ captures fishing-out effects—ideas get harder to find the further we advance.

The quantity of scientific work accomplished in this economy is

$$X_t(\sigma) \;=\; L_t(\sigma) \;+\; \frac{1}{\alpha_t(\sigma)}\, Q_t(\sigma).$$

Scientific work must be accomplished by scientists, whose supply is $H(s)$ for some skills $s \in \mathcal{S}_R$ that can be used in a subset of scientific work $\Xi(s)$. This implies

$$\sum_{\sigma\in\Xi(s)} L_t(\sigma) \;\leq\; H(s) \text{ for all } s \in \mathcal{S}_R.$$

On the other hand, completing scientific work with AI consumes computational resources. The *computational resource constraint* of the economy is now

$$\sum_{\sigma\in\Xi} Q_t(\sigma) + \sum_{\omega\in\Omega} Q_t(\omega) \;\leq\; Q_t.$$

The set of skills used for scientific work are assumed different from those used for production, denoted by $\mathcal{S}_P$.

To simplify the exposition, assume all scientific and production work are bottlenecks.

The two premises above are now strengthened as follows:

Premise 1′ (AGI for science). *For all work $\omega \in \Omega$ and scientific work $\sigma \in \Xi$, $\alpha_t(\omega)$ and $\alpha_t(\sigma)$ converge to some finite values $\alpha(\omega), \alpha(\sigma) \in (0, \infty)$ over time.*

I also strengthen Premise 2 to:

Premise 1′ (Exponential compute). *Q_t grows exponentially at rate $g_Q > 0$.*

The next proposition characterizes the limit behavior of the economy. To highlight key forces, I assume a constant fraction λ of compute is allocated to science. Naturally, there is some optimal value for λ that depends on how society discounts future consumption flows.

Proposition 9. *Suppose fraction $\lambda \in (0, 1)$ of compute allocated to science. All production and scientific bottlenecks are automated and output converges to*

$$Y_t \ = \ Z_t \ A \ \left((1-\lambda) \ Q_t \ + \sum_{s \in \mathcal{S}_R} CEU(s) \ H(s)\right), \tag{3}$$

where

$$Z_t \ \propto \ \left(B \ \left(\lambda \ Q_t \ + \sum_{s \in \mathcal{S}_P} CEU(s) \ H(s)\right)\right)^{1/\beta}. \tag{4}$$

As before, aggregate production possibilities in the economy become additive in compute and labor, with labor expressed in units of compute.

More novel, the stock of productive knowledge $Z(t)$ expands over time and scales with computing resources. In standard semi-endogenous growth models, $Z(t)$ scales with population, since scientific work requires human brains and these are limited by the size of the population. Here, by contrast, $Z(t)$ scales with the economy's computational resources Q_t because these can be used to automate all scientific bottlenecks.

The fact that AGI makes it feasible to complete scientific work with compute generates a compounded growth effect.

Proposition 10. *The growth rate of output and technology converge to*

$$g_Y \ = g_Q \ \left(\frac{1}{\beta} + 1\right) \qquad\qquad g_Z \ = g_Q \ \left(\frac{1}{\beta}\right)$$

The main difference with Proposition 3 is that, in an economy without other forms of scientific progress, output scales with compute and $g_Y \ = \ g_Q$. The possibility of carrying out science with compute implies that the growth rate is now higher by $g_Q \ \left(\frac{1}{\beta}\right)$. This is because some compute is used to expand productive knowledge, which then raises the productivity of *all* units of compute employed in production, generating increasing-returns in compute Q_t.

The above results clarify that, even in an AGI world where compute can be used for science, output scales with compute. There is not necessarily an intelligence or growth explosion, as this requires an exploding amount of compute.

One interesting aspect has to do with the prioritization of compute across uses.

Proposition 11. *Suppose the economy discounts future resources at a rate $\rho \geq 0$. In order to maximize welfare, in the long run, the economy allocates a fraction*

$$\lambda^* = \frac{1}{1+\beta+\rho}$$

of compute to science and the remaining to production. Along the transition, scientific bottlenecks are prioritized and $\lambda(t)$ starts above λ^*, eventually converging to λ^*.

This result provides some guidance on how an efficient economy organized to maximize discounted aggregate output would prioritize scarce compute.

In the short run, some priority should be given to scientific work, as this allows the rapid expansion of production knowledge, which makes the economy more productive at all future times. Other considerations not accounted for here, such as minimizing labor-market disruptions along the transition, call for even more prioritization of scientific work and a more gradual deployment of AGI for production (see for example Lehr and Restrepo, 2022).

In the long run, compute is allocated to both uses, and the economy eventually automates all production and scientific bottlenecks. The reason is that this strategy achieves the maximum scaling of output with compute. Using all compute for science yields a growth rate of

$$g_Y = g_Q \left(\frac{1}{\beta}\right).$$

Using all compute for production yields a growth rate of

$$g_Y = g_Q.$$

Using compute for both types of bottlenecks yields a higher growth rate because it exploits the increasing-returns to scale synergies between production and research introduced by scientific progress.

3 Discussion

This chapter analyzed the long-run implications of AGI for production, growth, and labor markets. The findings challenge both overly optimistic and overly pessimistic views about the future of human labor. On one hand, human work does not become obsolete. Because compute remains scarce, labor retains value by saving on computational resources. In addition, some accessory work may be left untouched by AGI, offering a stable, persistent set of roles for humans. On the other hand, while humans can still contribute economically, their role shrinks dramatically. The work that remains for people—whether bottleneck or accessory—commands a fixed value, bounded by the compute required to replicate it. In a growing economy, this means wages stagnate and the labor share collapses. Even skilled workers performing essential tasks will earn only what they save in compute—no more. In sum: humans can still work in an AGI economy—so long as they are endowed with time and skill—but their contribution becomes economically negligible. The economy keeps growing; we stay in place.

An interesting implication is the potential persistence of unautomated accessory work. For many socially intensive tasks—such as care work, hospitality, or therapy—the compute required to emulate human warmth or social intuition may be enormous. Even if such work is technically automatable, it may remain economically impractical to do so. As a result, these domains could continue to offer meaningful work for people. Yet these roles, while important, are not a source of growing income. The wages they offer are capped, possibly traded at a discount, and unresponsive to aggregate growth. Accessory work may provide continuity and stability in a transforming economy, but are not a source of growing riches providing human labor a central economic role.

The transition to AGI also presents distinct challenges. If the bottleneck is technological—i.e., if the key constraint is developing AGI systems themselves—then inequality may rise sharply during the transition, as certain types of labor temporarily gain extraordinary value. From the viewpoint of workers, the transition will feel uncertain and even unfair. People with some skills may see their wages fall rapidly to their compute-equivalent levels while others, by a sheer act of luck, may see their wages rise above it, earning a large premium just because the work they do turned out to be among the last we figured how to automate. This feels harder to navigate than a transition where compute binds, and wage growth is punctuated by gradual declines and the gradual need to reallocate labor. One crucial question here is how to help workers navigate a transition where technology binds. How can we help them share risks? Would other policies such as deploying the technology gradually help here (see also Lehr and Restrepo, 2022)?

As labor ceases to be the primary driver of value, economic policy must confront a basic question: how can we share the income generated by compute? In a world where AGI performs all bottleneck work, income generated by production flows to those who own or control computational resources. One approach is to redistribute the gains from compute through universal dividends. An alternative is to reimagine compute as a public or semi-public resource, akin to land or natural capital, with returns broadly shared.

Looking beyond wages and output, AGI also raises deeper questions about meaning and purpose. Historically, labor supplied the bottlenecks: it powered economic growth and gave individuals a central role in production. That connection is severed in the world of AGI. Human work no longer drives progress; it is no longer needed to improve living standards. If tomorrow half the population stopped working, no one would notice. In the AGI economy, we won't be missed.

Would people still choose to work, even when it makes no difference? Would work remain meaningful when it ceases to be economically essential? Or would we stop working altogether, not because we were replaced, but because we chose to step aside—seeking fulfillment in a world where our skills no longer matter?

References

Acemoglu, Daron, and David H. Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In *Handbook of Labor Economics*. Vol. 4, , ed. Orley Ashenfelter and David Card, 1043–1171. Amsterdam:Elsevier.

Acemoglu, Daron, and Pascual Restrepo. 2018. "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment." *American Economic Review*, 108(6): 1488–1542.

Acemoglu, Daron, and Pascual Restrepo. 2019. "Artificial Intelligence, Automation, and Work." In *The Economics of Artificial Intelligence: An Agenda*. , ed. Ajay Agrawal, Joshua Gans and Avi Goldfarb, 197–236. Chicago, IL:University of Chicago Press.

Aghion, Philippe, Benjamin F. Jones, and Charles I. Jones. 2019. "Artificial Intelligence and Economic Growth." In *The Economics of Artificial Intelligence: An Agenda*. , ed. Ajay Agrawal, Joshua Gans and Avi Goldfarb, 237–282. Chicago, IL:University of Chicago Press.

AI Impacts. 2023. "How much computing capacity exists in GPUs and TPUs in Q1 2023?" *https://wiki.aiimpacts.org/doku.php?id=ai_timelines:hardware_and_ai_timelines:computing_capacity_of_all_gpus_and_tpus*, Accessed: 2025-07-01.

Amodei, Dario. 2024. "Machines of Loving Grace." *Essay on Anthropic CEO's website*, https://www.darioamodei.com/essay/machines-of-loving-grace.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics*, 118(4): 1279–1333.

Bostrom, Nick. 2003. "Are You Living in a Computer Simulation?" *https://www.simulation-argument.com/*, Accessed: 2025-07-01.

Bostrom, Nick. 2014. *Superintelligence: Paths, Dangers, Strategies.* Oxford University Press.

Drago, Luke, and Rudolf Laine. 2025. "The Intelligence Curse: A Map to Maintaining Human Economic Relevance in the Age of AGI." *Essay series at intelligence-curse.ai*, Accessed July 2025.

Feenstra, Robert C., and Alan M. Taylor. 2017. *International Trade.* . 4th ed., New York:Macmillan Learning.

Hilbert, Martin, and Priscila López. 2011. "The World's Technological Capacity to Store, Communicate, and Compute Information." *Science*, 332(6025): 60–65.

Jones, Charles I. 1995. "R&D-Based Models of Economic Growth." *Journal of Political Economy*, 103(4): 759–784.

Kokotajlo, Daniel, Scott Alexander, Thomas Larsen, et al. 2025. "AI 2027: A Detailed, Falsifiable Forecast of the Next 2–3 Years." *AI Futures Project blog*, Accessed 2025-07-04.

Korinek, Anton, and Donghyun Suh. 2024. "Scenarios for the Transition to AGI." *arXiv preprint*, 2403.12107.

Kortum, Samuel. 1997. "Research, Patenting, and Technological Change." *Econometrica*, 65(6): 1389–1419.

Lehr, Nils H., and Pascual Restrepo. 2022. "Optimal Gradualism." National Bureau of Economic Research NBER Working Paper 30755.

Open Philantropy. 2020. "How Much Computational Power Does It Take to Match the Human Brain?" https://www.openphilanthropy.org/research/how-much-computational-power-does-it-take-to-match-the-human-brain/, Accessed: 2025-07-01.

Sandberg, Anders, and Nick Bostrom. 2008. "Whole Brain Emulation: A Roadmap." Future of Humanity Institute, University of Oxford Technical Report 2008-3.

Segerstrom, Paul S. 1998. "Endogenous Growth without Scale Effects." *American Economic Review*, 88(5): 1290–1310.

Trammell, Philip, and Anton Korinek. 2023. "Economic Growth under Transformative AI." National Bureau of Economic Research NBER Working Paper 31815.