We Won't be Missed: Work and Growth in the AGI World*

Pascual Restrepo
pascual.restrepo@yale.edu
Yale University

October 6, 2025

Abstract

This chapter explores 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 supplementary work—tasks essential vs. non-essential for unhindered growth. As computational resources expand: (i) the economy automates all bottleneck work, (ii) some supplementary work may be left 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 article was prepared for the NBER conference volume on The Economics of Transformative AI. I thank my discussants, Neil Thompson, Danial Lashkari, and Omeed Maghzian, for helpful comments. Disclosure: Since May 9th, 2025, I have been part of Anthropic's Economic Advisory Council. For slides and an online supplement with derivations, visit https://campuspress.yale.edu/pascualrestrepo/.

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 perform their tasks.

The economic problem in the AGI world is how to allocate finite (but growing) computational resources and human skill to accomplish work needed for output. The chapter introduces a distinction between bottleneck and supplementary work:

- *Bottleneck work* comprises tasks essential for economic growth. Output cannot expand indefinitely unless inputs in bottleneck tasks also expand or become infinitely valuable.

- *Supplementary work* is non-essential for growth. Output can expand indefinitely even if the inputs in these tasks remain fixed.

My primary theoretical result is that all bottleneck work will eventually be automated, with most of it carried by autonomous AI systems. Once this occurs, production shifts from a multiplicative relationship between compute (the stock of computational resources) and human skill to an additive one, and the long-run growth rate of the economy depends on the growth rate of compute.

Even after AGI performs all bottleneck tasks, people may still work and earn positive wages. Some skills may be used in bottleneck work, where their contribution is valued by the amount of compute they save. The existence of AI systems capable of performing this work does not mean wages fall to zero, since there is a real opportunity cost to using compute instead of people. Other skills may be used for supplementary tasks left to humans and not automated. This work also commands a positive wage, but it is capped by the compute cost required to automate it.

AGI fundamentally changes the role of labor and how it is valued. Before AGI, human skill was the main driver of output, and wages reflected the scarcity of skills needed for bottleneck tasks. In an AGI world, compute takes that central role, and wages are anchored to the computing cost of replicating human skill. While human wages remain positive—and on average exceed those in the pre-AGI world—their value becomes decoupled from GDP, the labor share converges to zero, and most income eventually accrues to compute.

In addition, I expand the analysis to an economy where AGI can be used for scientific work, accelerating the pace of technological progress. Without AGI in 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 a shrinking population, but it does not lead to a singularity or an infinite growth explosion.

The analysis complements existing work on the economics of AI, including contributions in a previous NBER volume (see Aghion, Jones and Jones, 2019; Acemoglu and Restrepo, 2019). I build on task models, which study how to accomplish work tasks using workers of different skill or automated technologies (see Autor, Levy and Murnane, 2003; Acemoglu and Autor, 2011; Acemoglu and Restrepo, 2018). As in these models, the goal is to understand how the economy allocates work between human labor and AGI systems, and derive the implications of this allocation for wages and output. The focus on the transition to AGI is shared with complementary work by Anton Korinek and collaborators (see Trammell and Korinek, 2023; Korinek and Suh, 2024). Beyond economics, there is also a large and growing literature of world-building exercises imagining how AGI transforms the economy and society (see Kokotajlo et al., 2025; Drago and Laine, 2025).

The chapter is organized as follows. Section 1 analyzes a production economy where AGI can be used to complete all production work with compute. Section 2 extends the analysis to a semi-endogenous growth setting where AGI is applied to scientific work. Section 3 provides some interpretation for my findings.

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 specific types of work indexed by ω. Ω is a large discrete set, listing all forms of productive work.

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

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

This work can be accomplished by human labor $L_t(\omega)$ or using computational resources $Q_t(\omega)$. The quantity of work produced by people per unit of time is normalized to 1. Emulating or replicating this work using computers costs $\alpha_t(\omega)$ units of compute. If $\alpha_t(\omega) = \infty$, it is impossible to replicate the human skill needed to complete ω with computers.

The quantity of output produced is

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

where F is increasing, smooth, differentiable, concave, and exhibits constant returns to scale.

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 portability of skills. If $\Omega(s)$ is a one-to-one mapping from $\mathcal{S}$ to Ω, skills are work-specific. If $\Omega(s) = \Omega$, there is a single universal skill.

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 model features two forms of technological progress: algorithmic progress (reductions in $\alpha_t(\omega)$, possibly from infinity to finite values) and advances in computing capabilities (the expansion of Q_t over time). The goal is to characterize the behavior of an economy that (i) develops *Artificial General Intelligence* and (ii) has enough computational resources to run it.

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. Developing 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.[1]

[1]The definition used here differs from other concepts, such as *Transformative AI* (defined as "AI developments that

While this premise is the point of departure in my analysis, it is helpful to clarify three points about what is being assumed.

- **Feasible vs Realized:** The definition of AGI states that, in principle, we can replicate what people do if we throw in enough compute. AGI thus makes it feasible to accomplish all work with compute, but this does not mean it is practical or efficient to automate everything. Depending on the computing costs $\alpha(\omega)$, it may be better to leave some tasks to humans and allocate our finite computational resources elsewhere.

- **Physical work:** Some work requires interacting physically with the world. The premise here is that, when needed, computer systems can control machines 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 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 (from a functional point of view), is at your disposal at any time, knows you perfectly, never gets tired, and is open to use 100% of its capabilities to help you overcome your anxieties. The system may be too costly to be practical, 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 sick children from accessing this medical attention because you insist it 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.

significantly transform the economy") or *Super-intelligence* (AGI that recursively self-improves itself, reaching levels of intelligence that far exceed human ones).

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

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

This 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.[2]

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.

Bottlenecks: In what follows, I let $F(\{X(\omega)\})$ denote the output obtained when the quantity of work is $X(\omega)$ and $F_{\omega_0}(\{X(\omega)\}) > 0$ denote the marginal gain from an additional unit of $X(\omega_0)$.

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

Definition 1. *Work ω_0 is bottleneck if, for any $\{X_t(\omega)\}$ such that $F(\{X_t(\omega)\})$ is unbounded, either i. $X_t(\omega_0)$ is unbounded or ii. $F_{\omega_0}(\{X_t(\omega)\})$ 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.

[2]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.

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 and lacks any of these components.

Definition 2. *Work ω_0 is supplementary if, there is $X_t(\omega)$ such that $F(\{X_t(\omega)\})$ is unbounded while both $X_t(\omega_0)$ and $F_{\omega_0}(\{X_t(\omega)\})$ are bounded.*

Supplementary work is the opposite of a bottleneck: the economy can keep growing while this type of work remains fixed, without its price rising excessively. Potential examples include arts and crafts, literature, hospitality, design, customer support, and judicial work. Even the work of academic economists may prove supplementary, as it is unlikely to become ever more valuable.

What these examples share is that one can imagine a future where people view these forms of work as valuable and desirable, but would not view a reduction in their supply as an absolute crisis. These are, of course, only illustrative. The boundary between bottleneck and supplementary work will depend on future preferences, the structure of production, the set of available alternatives, and the existential problems, challenges, and attitudes of 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 supplementary work may be left to labor.*

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

This last result seems to counter Ricardo's principle of *comparative advantage*. The AGI economy is a world where we engage in trade with "a country of geniuses in a data center," exchanging compute for work.[3] Wouldn't Ricardo's principle imply that the AGI and human countries should specialize to maximize the gains from trade? Why is it optimal for AGI to produce all bottlenecks?

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 while human labor specializes in the remaining components. Over time, the economy becomes unbalanced, producing an expanding quantity of the first type of work and a fixed one of the later. This imbalance cannot be optimal: to keep the economy growing one must expand *all* bottleneck work at the same rate. This imbalance is efficient only if non-automated work is supplementary, as claimed in the proposition.[4]

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

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

This gives the computational resources needed to replicate work carried by people of skill s in the most computationally complex task they perform. Premise 1 implies $\mathrm{CEU}(s)$ is finite for all s.

In the long run, skills fall in two groups. $\mathcal{A}$ represents skills such that some of the work in $\Omega^*(s) = \arg\max_{\omega \in \Omega(s)} \alpha(\omega)$ is automated. These skills specialize in $\Omega^*(s)$ and compete with AI systems in the production of this work. $\mathcal{N}$ represents skills such that work in $\Omega^*(s)$ is supplementary and not automated. These skills produce supplementary non-automated work, automated work outside $\Omega^*(s)$, or both. All supplementary non-automated work is produced by skills $s \in \mathcal{N}$.

Proposition 2. *Aggregate production converges to*

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

where $A > 0$ is the marginal rate of transformation of compute into output and N a constant returns to scale

[3] This metaphor is from Anthropic's Amodei (2024).

[4] The idea that Ricardo's principle of comparative advantage calls for *full* specialization is wrong. 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 pinning their relative prices.

and increasing function, bounded above by

$$N(\{H(s)\}_{s\in\mathcal{N}}) \;<\; A \sum_{s\in\mathcal{N}} CEU(s)\, H(s).$$

Aggregate production possibilities in the economy become additive in compute and human skill, with skills in $\mathcal{A}$ 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 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.

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 in all existing work. We can write output (given an allocation of labor and assuming $\alpha_t(\omega) = \alpha(\omega)$) as

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

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

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

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

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

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

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

These steps clarify the economics behind Proposition 2. AGI allows us to produce all bottleneck work with compute. At that 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 people 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).

With supplementary work, output continues to be linear in compute, the compute-equivalent units of skills $s \in \mathcal{A}$, and the contribution of skills used for non-automated work, summarized by the N function. This contribution is bounded above by the compute-equivalent units of all skills $s \in \mathcal{N}$. Otherwise, some of this work would be automated, since the benefit would exceed the cost.

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 with AGI, the economy can grow simply by expanding computational capabilities. No other form of technological progress is required for sustained growth. This is true even if there is supplementary work left unautomated, as this work does not hinder growth.

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

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

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

for some $0 < \widetilde{CEU}(s) < CEU(s)$.

Imagine first that AI systems produce all work in Ω. The proposition shows that people would still be paid the value of the compute needed to replicate their skills. This highlights an important distinction: AGI does not render labor redundant; it makes it replicable through costly computation, permanently altering how it is valued. Human skill remains valuable because it accomplishes useful work, saving scarce computational resources.

Suppose now that there is supplementary work that is not automated and performed by workers of skill $s \in \mathcal{N}$. Can such supplementary work provide a source of growing wages? No. The proposition shows that people performing supplementary work that is not automated earn wages below A CEU(s). Otherwise, we would be better off automating such work. The reason why this work is left unautomated is that we already have too many workers to do it, causing wages to be too low to justify the use of our scarce computational resources in the first place.

The significant result is that wages decouple from output growth, which in the AGI economy is driven entirely by expanding computing resources. Labor income remains bounded by the compute required to replicate it, so as the economy expands, 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 is bounded by the share of human compute in our total computational resources, measured by

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

This converges to zero over time. The numbers on computational resources cited above give a sense of magnitudes. The value of human compute, $\sum_s \mathrm{CEU}(s)\, H(s)$, 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 is paid its compute-equivalent value, the labor share plummets, as total compute is projected to far exceed human compute. Most income will accrue to owners of computing resources, which also account for most (or virtually all) of the output produced.

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 poorer. 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 increases the value of human labor as a whole.

Proposition 6. *Consider a pre-AGI economy with $Q_0 = 0$ and $\alpha_0(\omega) = \infty$ at $t = 0$, transitioning to an AGI* *world thereafter. For all $t \geq 0$, the sum of wages exceeds those paid in the pre-AGI economy*

$$W_t \;=\; \sum_s W_t(s)\, H(s) \geq \underbrace{\sum_s W_0(s)\, H(s)}_{W_{pre\text{-}AGI}} .$$

The proposition shows that workers *as a group* benefit from transitioning to the AGI economy. The arrival of AGI cannot make us collectively worse off, since we could always set up a no-AGI zone and carry our lives as in the pre-AGI world. 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 at every point in the transition to AGI. There is no rearrangement that can raise output. Suppose we take a representative group consisting of a small fraction $n > 0$ of the workforce and assign them to a no-AGI zone, where they replicate the way production was carried out before the transition. These workers would produce $W_{\text{pre-AGI}} \times n$ units of output in the no-AGI zone (an implication of constant-returns to scale), but this would also result in a loss of output $W_t \times n$, equal to their marginal product in the post-AGI world. Because this rearrangement cannot raise output, the cost exceeds the gains: $W_t \geq W_0$. Even though wages are capped by the value of compute needed to replicate human skill, the proposition shows that total labor income in the post-AGI world exceeds wages in the pre-AGI economy.

The argument above assumes that compute and human skill are the only scarce resources limiting output. If the economy also relies on other finite resources, the logic breaks down, since any no-AI zone must compete with the AGI economy for these inputs. Still, the scope for AGI to make human workers collectively less valuable is limited, in part because AGI also helps expand the supply of other potential limiting factors, such as land and energy.[5]

1.2 Transition to AGI

The path from today's economy to the AGI world depends on the relative pace of compute expansion (Q_t) and algorithmic progress (the α's). Two polar cases illustrate the possibilities.

[5]Even if it did not, the argument above shows that non-compute income—the sum of wages and returns to land and other scarce factors—would still rise in an AGI world. Even if we ignore income from compute entirely, AGI cannot make us collectively worse off; we own these other factors just as we own labor.

- **Algorithm-binding transition.** Work ω is automated immediately once it becomes feasible. This corresponds to a world where compute is plentiful.

- **Compute-binding transition.** Work ω is automated only later, once the economy accumulates enough compute to make automation worthwhile.

This section traces how these two transitions unfold. The actual path could combine features of both: at some points algorithmic progress binds, while at others compute does.

For clarity, I work in continuous time, assume all work is a bottleneck, and represent algorithmic progress by $\alpha_t(\omega)$ switching from infinity to a finite limit $\alpha(\omega)$ at date $T(\omega)$. Tasks in $\Omega(s)$ are assumed to have different computational requirements, so they can be ranked in this dimension. Finally, I assume that holding $X_t(\omega_0)$ constant, $F_{\omega_0}(\{X_t(\omega)\})$ increases as compute Q_t expands and is used for other work. This condition rules out cases of de-automation and simplifies the analysis.

Proposition 7. *Suppose compute binds at all times. Along the transition:*

- *The marginal value of compute A_t decreases continuously and converges to $A > 0$, while total wages W_t increase continuously during the transition and converge to $\sum_s CEU(s)\ H(s)$.*

- *For all skills s, wages $W_t(s)$ follow continuous paths.*

- *The work in $\Omega(s)$ is automated from least to highest computational requirements:*

  1. *The automation of all $\omega \in \Omega(s)$ with $\alpha(\omega) < CEU(s)$ occurs first and causes wages $W_t(s)$ and employment $L_t(\omega)$ to decrease continuously for an interval of time, with wages decreasing at the same rate as A_t during this interval, until $L_t(\omega) = 0$.*
  2. *The automation of all $\omega \in \Omega(s)$ with $\alpha(\omega) = CEU(s)$ happens last, only after wages have reached a level $W_t(s) = A_t\ CEU(s)$, and keeps wages at this level from there on.*

When compute binds, the transition to the AGI limit is smooth and gradual, featuring a growing path for total wages W_t. For every skill s, wages follow a continuous path, punctuated by temporary episodes when some of their tasks are automated, as shown in Figure 1. During these episodes, workers gradually reallocate away from the automated work, which is then fully taken by AGI. This process continues until only the most computationally complex work with $\alpha(\omega) = \text{CEU}(s)$ remains (ω_3 in the Figure). This work is automated when wages equal the cost of producing it with compute, and thereafter wages equal A_t CEU(s) and converge to the limit values in Proposition 4. Once this happens, workers do not reallocate and continue earning their equivalent in compute.

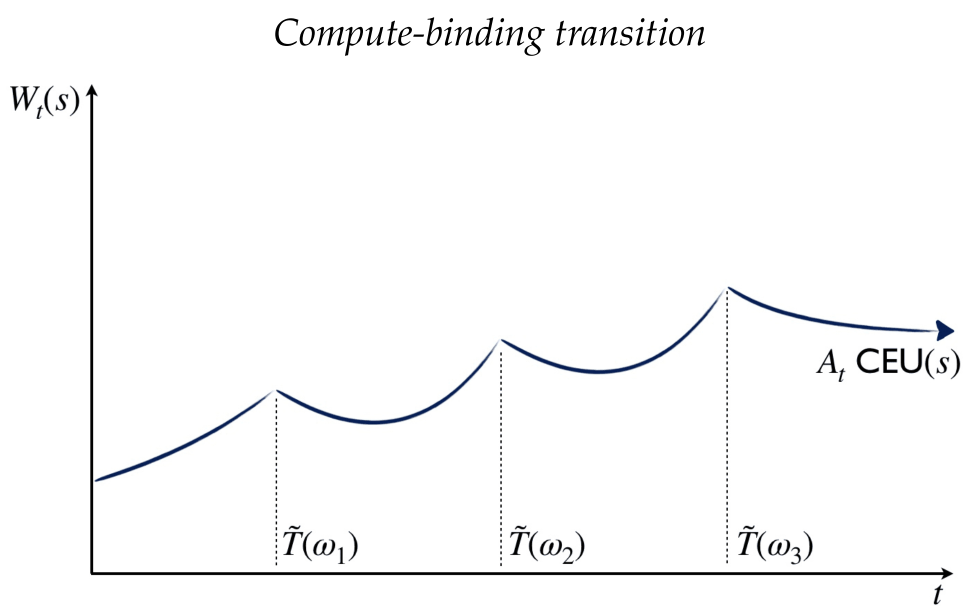

Figure 1: Continuous time path for wages along a compute-binding transition. The figure shows an example for a skill performing work $\Omega(s) = \{\omega_1, \omega_2, \omega_3\}$ automated at dates $\tilde{T}(\omega_1) < \tilde{T}(\omega_2) < \tilde{T}(\omega_3)$.

Proposition 8. *Suppose algorithms bind at all times. 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, while total wages W_t may feature negative or positive jumps at points, before eventually converging to $\sum_s CEU(s)\,H(s)$.*
- *For all skills s, wages $W_t(s)$ follow discontinuous paths.*
- *The work in $\Omega(s)$ is automated at time $T(\omega)$, as dictated by algorithmic progress. Let $T(s) = \max_{\omega \in \Omega(s)} T(\omega)$. Then:*
  1. *The automation of all $\omega \in \Omega(s)$ with $T(\omega) < T(s)$ causes employment $L_t(\omega)$ to jumps down while wages $W_t(s)$ can jump up or down (if the productivity gains are small).*
  2. *The automation of the last $\omega \in \Omega(s)$ at $T(s)$ causes wages to jump down to $W_t(s) = A_t\,CEU(s)$, which then converges monotonically to their long-run value.*

When algorithmic progress binds, the transition is jagged, with sudden income losses for some and gains for others. The proposition describes a world where compute is abundant, firms experi-

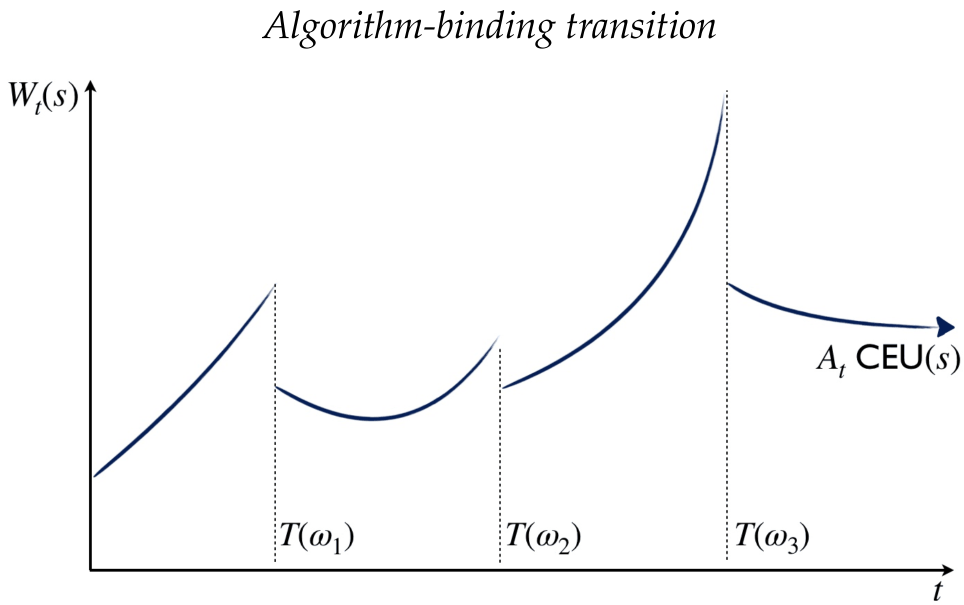

Figure 2: Discontinuous time paths for wages along an algorithm-binding transition. The figure shows an example for a skill with $\Omega(s) = \{\omega_1, \omega_2, \omega_3\}$ automated at dates $T(\omega_1) < T(\omega_2) < T(\omega_3)$.

ment in the background, and when an application succeeds, it is adopted immediately, disrupting the labor market. Employment in the newly automated tasks drops at once, and wages for workers with those skills may fall if the productivity gains are small (in the terminology of Acemoglu and Restrepo, 2018). This scenario is illustrated in Figure 2.

The proposition also shows that automating the last remaining task in $\Omega(s)$ necessarily causes a wage drop for skill s, from some level $W_t(s)$ down to $A_t\text{CEU}(s)$, since workers have no other tasks to reallocate to. The pre-automation premium reflects the scarcity of skill s, as it was the only way to produce bottleneck work that we wished to automate but could not. Once all tasks in $\Omega(s)$ are automated, this premium vanishes and wages fall to their compute-equivalent level.

2 Scientific Work

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 in the AGI economy is

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

Scientific work can be carried by scientists, with skills $s \in \mathcal{S}_R$, and in quantity $H(s)$. Scientists of skill s can perform 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 above by $\mathcal{S}$.

For clarity, I 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 2′ (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 a fraction $\lambda \in (0, 1)$ of compute allocated to science. All production and scientific bottlenecks are automated and output converges to*

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

where

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

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, $Z(t)$ 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 we discount future utility at a rate $\rho \geq 0$ and trade it off with a constant intertemporal elasticity of substitution $1/\gamma$. In order to maximize welfare, the economy allocates a fraction*

$$\lambda^* = \frac{g_Q}{\beta\rho + (1+\beta)\gamma g_Q}$$

of compute to science and the remaining to production in the long run. Moreover, if the economy starts below the scientific frontier (given its research resources), scientific bottlenecks should be prioritized during the transition, and $\lambda(t)$ converges to λ^* from above.

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 scientific knowledge. 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 scientific progress.

3 Discussion

This chapter examined the long-run implications of AGI for production, growth, and labor markets. AGI is algorithmic progress that makes all economically valuable work feasible with compute. With sufficient computational resources, bottleneck tasks are fully automated, while some supplementary work may remain in human hands.

The findings challenge both overly optimistic and pessimistic views of the future of human labor.

On one hand, supplementary work may provide stable roles for humans. The model opens up the intriguing possibility that much of today's work may not be essential for future growth and may never be automated. Instead, compute may be directed toward bottleneck work critical for future progress—such as reducing existential risks, defending against asteroids, or mastering fusion energy—leaving large parts of the labor market unchanged. Socially intensive work—such as hospitality, live performances, and entertainment—may be non-essential for future growth, costly to replicate with compute, and thus remain in human hands. These domains could continue to offer familiar and meaningful work.

More fundamentally, AGI does not render human skills obsolete; it revalues them. Because compute is scarce, skills are valued at the opportunity cost of compute required to replicate them. In fact, if compute and human skill are the only scarce resources, average wages are higher in a post-AGI world.

On the other hand, labor's relative role shrinks. Once bottlenecks are automated, the value of work—whether bottleneck or supplementary—is bounded by the compute required to replicate it. Wages become decoupled from output, and the labor share collapses. Skilled workers performing essential bottlenecks earn only what they save in compute. Supplementary work may offer continuity but not rising wages.

The type of transition matters. When compute is the binding constraint, adjustment is gradual, with workers slowly reallocating as tasks become automated. When algorithmic progress is the constraint, the transition is jagged, uncertain, and risky. Inequality may rise sharply: workers whose tasks cannot yet be automated enjoy large temporary wage premiums, while others face sudden wage declines as theirs are. A central policy question is how to help workers share these risks and navigate a jagged transition (see also Lehr and Restrepo, 2022; Restrepo, 2025).

Distribution is another central issue. In an AGI economy, most income accrues to owners of compute. One approach is to redistribute these gains through universal income. Another is to treat compute as a public resource—akin to land or natural capital—and distribute its returns broadly.

Finally, AGI raises questions of meaning and purpose. Historically, work provided not only income but also recognition that one's efforts improved society's well-being. Work gave people the sense that they would be missed. In an AGI world, that connection is severed. Human skill is no longer needed to improve living standards in an appreciable way. Today, if half of us stopped working, the economy would collapse. In the AGI world, we would not be missed.

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.

Restrepo, Pascual. 2025. "Policy for a Changing Landscape." In *Steering Structural Change: Rethinking Government Policy to Support Fundamental Economic Transformations*. , ed. Pierre-Olivier Gourinchasa, Maurice Obstfeld and Petia Topalova, 241–264. Cambridge, MA:MIT Press.

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.