AI and Our Economic Future

Charles I. Jones*

Stanford GSB and NBER

May 18, 2026 — Version 2.0

In preparation for the Journal of Economic Perspectives

Abstract

Artificial intelligence (AI) will likely be the most transformative technology of the modern era. What if machines — AI for cognitive tasks and AI plus advanced robots for physical tasks — can perform every task a human can? This essay makes three main points. First, even though US growth rates have been stable at roughly 2 percent per year for 150 years, it is distinctly possible that automating intelligence leads economic growth rates to accelerate. Second, this acceleration is likely to be slowed by the presence of "weak links." While we each have access to 100 million times more transistors on our desktop computer than people in the 1970s, we are not 100 million times more productive. Computers can invert matrices at lightning speed, but we humans must still decide what matrix to invert, what hypothesis to test, etc. Accelerating economic growth requires the vast majority of the weak links to be automated away, which delays the large gains. Finally, even though weak links slow the benefits, they may actually speed up the risks. When a chain is only as strong as its weakest link, damaging one link in the chain can be very costly. A powerful AI that is superhuman at software engineering could be misused by a bad actor to do substantial harm by hacking the financial system or a virology lab.

*Charles I. Jones is the Stanco25 Professor of Economics, Stanford Graduate School of Business, Stanford, California. He is also a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. His email address is chad.jones@stanford.edu. The author is grateful for helpful comments from David Autor, Basil Halperin, Ben Jones, Tom Houlden, Pete Klenow, Jeffrey Kling, Narayana Kocherlakota, Jonathan Parker, Tim Taylor, Chris Tonetti, Phil Trammell, and Heidi Williams.

1. Introduction

Artificial intelligence (AI) will likely be the most transformative technology of the modern era. Earlier technologies such as electricity, semiconductors, and the internet profoundly reshaped economic activity and increased living standards throughout the world. In some sense, AI is simply the latest in this line of general purpose technologies and at a minimum should continue the economic transformation that has been ongoing for the past century.

However, a case can be made that this time is different. Automating intelligence itself — that is, creating a technology that can perform any cognitive task that humans can do today — would have broad ramifications. Previous technological advances replaced humans in many physical tasks or in a narrow set of cognitive tasks, making human intelligence more valuable. But what if machines — AI for cognitive tasks and AI plus advanced robots for physical tasks — can perform every task a human can, at least as cheaply? While this outcome is by no means guaranteed in the next several decades, a number of observers expect that it will occur (Amodei, 2024; Aschenbrenner, 2024; Brynjolfsson et al., 2025; Kokotajlo et al., 2025).

What does economics have to say about this possibility, and what might our economic future look like?

We begin by outlining two scenarios for the impact of artificial intelligence on the economy: one in which AI drastically accelerates economic growth and another in which AI is "business as usual." Both scenarios are possible, and both scenarios have features that are informative about the future consequences of AI.

The future probably lies somewhere in between these two scenarios, but where? To shed insight on this question, we describe task-based models of economic growth in which tasks are complementary — a "weak link" framework. Just as a chain is only as strong as its weakest link, the economy may be limited by whatever tasks are not yet automated — that is, by tasks which are performed by slowly-improving humans rather than by rapidly-improving machines. This feature tends to slow the transformative effects of AI. But "slow" does not mean "eliminate," and AI is likely to dramatically transform the economy over a time horizon measured in decades. In the United States, per capita GDP growth has averaged 2 percent per year for the past 150 years and been remarkably stable. Nevertheless, if AI eventually automates away nearly all the weak links in the economy, economic growth could accelerate significantly, with rates potentially exceeding 5 percent per year (Jones and Tonetti, 2026).

Powerful and transformative AI also entails risks; it is a double-edged sword. What happens to labor markets, inequality, and the political equilibrium when machines can perform nearly every task that humans can? More speculatively, transformative AI enables potentially catastrophic risks, for example through cyberattacks, biological vectors, and autonomous weapons. As the leading AI labs have emphasized, these possibilities deserve serious consideration.

It is surely obvious but also important to acknowledge that no one knows what the consequences of automating intelligence will be. This essay provides one perspective based on economic research, but the perspective is inherently speculative and will likely prove to be incorrect on many dimensions. Nevertheless, I have found that the tools of economics are helpful in providing guideposts and insights as we consider how AI will impact our economic future.

2. Two Scenarios for the Future of Growth

To begin, let me sketch out two possible futures for the macroeconomy. These scenarios are intentionally extreme, at both ends of the spectrum. In the first, artificial intelligence has massive economic impacts and dramatically accelerates economic growth. In the second, AI is more "business as usual" or a "normal technology" to use the phrase coined by Narayanan and Kapoor (2025). Both scenarios are possible outcomes in my view; I'll explain later how I think of where in between these two outcomes our future may lie.

Each scenario will describe changes that could occur over 25 or 30 years. Thus, the time period I have in mind is roughly one generation: what might the world look like as a child born today grows into an adult, or as a young adult now entering the workforce moves though their career?

2.1 AI Accelerates Economic Growth

What would the next several decades be like if artificial intelligence accelerates economic growth and has truly profound macroeconomic consequences? Scenarios along these lines have been presented by various authors.[1] While the scenarios are inherently speculative, I find two aspects to be intriguing. First, there is already evidence for their early stages, and second, each step does not seems impossible.

The scenarios typically begin with AI raising the productivity of software engineers, which is a process that is already solidly under way. As one of a parade of similar examples, when Anthropic introduced Claude Opus 4.5 in November 2025, the firm highlighted Claude's exceptional performance on a two-hour take-home exam they give to prospective software engineering hires: the AI model scored higher than any human candidate ever (Anthropic, 2025).

Epoch AI estimates that the amount of "effective compute" (that is, total computing power adjusted for the quality of algorithms and software) used to train artificial intelligence models is rising annually by a factor of ten: a factor of 4 from more and better computer chips and a factor of 2.5 from better algorithms (Ho et al., 2024; Epoch AI, 2025a). These investments result in rapid improvements: for example, METR (2026) has developed a test of the extent to which AI tools can autonomously carry out programming projects. In 2020, ChatGPT 3 could complete a task that would take a human 9 seconds with a 50 percent probability of success. By June 2024, AI could solve an 11-minute project with this accuracy. Two years later in mid-2026, the corresponding 50 percent success time had risen to 12 hours. The METR time horizon for software engineering has been doubling roughly every 5 to 7 months. There is some suggestion that progress has accelerated, with the doubling time falling to just 4 months since the reasoning models (e.g. o1-preview) were introduced in late 2024 (Emberson, 2025).

One of the frontier uses of artificial intelligence models is creating AI agents, models adept at performing tasks currently done by humans on computers. Beyond coding, these include writing and editing documents, synthesizing reports, building spreadsheet models, developing slide presentations, making videos, brainstorming ideas, suggesting science experiments, and proving theorems in mathematics. In many fields and for many cognitive tasks, AI already performs at the level of a competent graduate student. Given the rapid progress over the past five years, AI models a decade from now may be able to accomplish nearly any of the tasks that talented humans can currently accomplish on computers.

[1] For example, see Davidson (2021), Kokotajlo, Alexander, Larsen, Lifland, and Dean (2025), Epoch AI (2025b), Cunningham (2025), and Davidson, Halperin, Houlden, and Korinek (2026).

One such task is doing AI research. AI models may themselves, with no human involvement, eventually be able to create better AI models, a phenomenon known as *recursive self-improvement.* Combining this flywheel effect with rapid increases in computing power, it is possible that sometime in the next decade — and perhaps much sooner — we will have access to what Dario Amodei, CEO of Anthropic, has called "a country of geniuses in a data center" (Amodei, 2024).

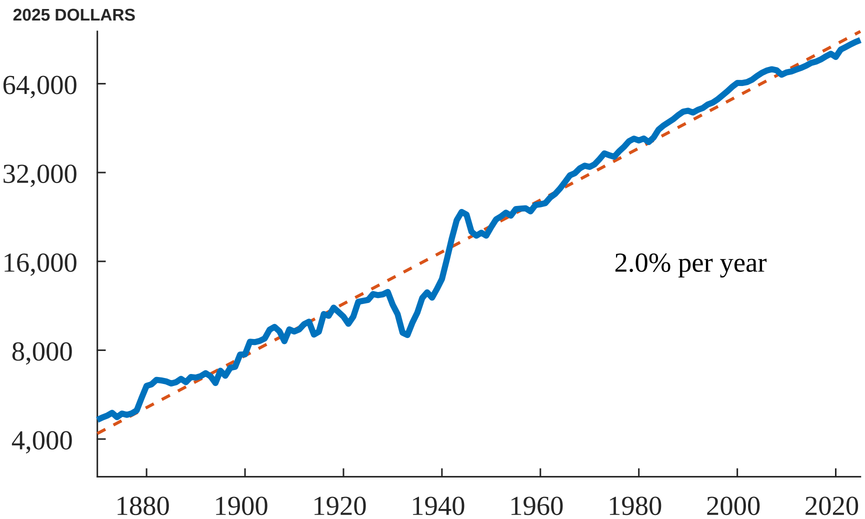
Millions of virtual geniuses-in-a-data-center, each running much faster than the speed of the human mind, could accelerate the discovery of new ideas. We've already seen new discoveries by AI models: one vivid example is AlphaFold, the model that "solved" the problem of how to determine the three-dimensional structure of more than 200 million proteins given only the sequence of amino acids that defines them, leading to the 2024 Nobel Prize in chemistry for Demis Hassabis and John Jumper of Google DeepMind. An economy of AI geniuses could design new pharmaceuticals and predict which drugs are likely to do best in clinical trials with minimal side effects. They could synthesize the existing research literature and advise scientists on what lab experiments to run in fields ranging from biochemistry to materials science to nuclear fusion and clean energy.

In the longer run, genius-level AI models could use virtual simulations and human-run experiments to design and build better robots, iterating until exceptional performance is achieved. Eventually, this flywheel could result in AI models that can perform nearly all cognitive tasks and robots that are designed, produced, and managed by AI that can perform nearly all physical tasks. If this scenario were to come to pass, even over the next half century, the gains in productivity and living standards would surely be tremendous. In standard growth models, this outcome would produce accelerating economic growth; in some cases, this acceleration could lead to growth rates exceeding 10% per year or more (Aghion, Jones, and Jones, 2019; Jones and Tonetti, 2026).

2.2 AI as "Business as Usual"

The US economy in the last half century has experienced dramatic increases in computing power, as illustrated by Moore's Law. Back in 1965, Gordon Moore, who was then director of research at Fairchild Semiconductor, wrote a paper predicting that the density of transistors on a computer chip would double every two years, a prediction

Figure 1: Average US Income per Person



Note: Data from U.S. Bureau of Economic Analysis (2025) since 1929 and from Barro and Ursua (2010) for 1870 to 1928.

that has largely been borne out across six decades (Moore, 1965). The dramatic fall in the price of computing has brought technological miracles like the internet, but has not led to a sustained boost in the US growth rate. Is there also a scenario in which artificial intelligence leads to "business as usual" for future growth?

Figure 1 displays my favorite graph in economics: real GDP per person in the United States for the past 150 years. On a log scale, the time series is roughly a straight line. Indeed, Moore's Law and US GDP per capita growth are in some sense duals, illustrating the "rule of 70" that the product of the growth rate and doubling time under constant exponential growth equals 70: transistor density doubles every two years and grows at 35 percent per year while per capita US GDP grows at 2 percent per year and doubles every 35 years,

Consider the astounding innovations that underlie the graph. In the 1870s, Thomas Edison's experiments with electric lighting were just getting underway. Fifty years later, electrification had transformed the economy, both in factories and in city life. Throughout the 150 years, innovations such as the internal combustion engine, airplanes, vacuum tubes, antibiotics, transistors, semiconductors, personal computers, and the internet profoundly changed living standards. Many of these innovations are what economic historians call "general purpose technologies," whose transformative effects extend throughout the economy. Many also automated some of the tasks involved in creating new ideas — say, through improvements in scientific tools and equipment – raising the productivity of research and idea generation. Yet apparently none of these innovations changed the long-run growth rate of the US economy.

How can we understand this disconnect? One natural hypothesis is that within any technology field, ideas get harder to find (Bloom, Jones, Van Reenen, and Webb, 2020). The steam engine runs out of steam. Without the discovery of the next general purpose technology, one might expect economic growth to slow down. From this perspective, each of these new general purpose technologies did indeed raise the growth rate of the economy in a counterfactual sense: but for their invention, the counterfactual is that growth would have slowed. The continued emergence of amazing new technology classes is what made sustained growth at 2 percent per year possible. Perhaps artificial intelligence is just the latest general purpose technology that lets 2 percent growth continue for several more decades. Notice that even in this scenario, AI potentially has large transformative effects, because the counterfactual is one in which growth would otherwise have slowed.

Another complementary lesson from economic history is that it can take decades for a new innovation to diffuse throughout the economy. David (1990) first made this point in the context of the steam engine and the electricity-generating dynamo. Robert Solow famously quipped in 1987: "You can see the computer age everywhere but in the productivity statistics" (Solow, 1987). Profound new technologies require many complementary innovations before they can take full effect. In the case of electricity, factories that had been driven by steam power needed to reorganize and even relocate production before they could realize the full productivity gains from small electric motors that could be placed throughout the factory. Organizational, product, and process changes needed to be implemented to take advantage of information technology (Brynjolfsson and Hitt, 2000). For the new artificial intelligence, the lesson from economic history is that the effects on GDP and productivity may take decades to realize their full transformative effects.

2.3 Comparing the Scenarios

Both of the scenarios I've laid out have merit. The "accelerating growth" scenario is based on direct evidence about the rapidly evolving capabilities of artificial intelligence. Theoretical models of economic growth in which AI automates most of the tasks in the economy can formally produce explosive economic growth (Aghion, Jones, and Jones, 2019). On the other hand, automation of production has been ongoing for more than 200 years, and transformative innovations such as electricity, semiconductors, and the internet have coexisted with remarkably stable economic growth at 2 percent per year.

In either scenario, the effects of artificial intelligence are large and profound, just as they were with electricity and the internet. Still, exactly how large and how profound is far from clear. Moreover, the future is likely to lie somewhere in between; substantial uncertainty surrounds the future effects of artificial intelligence on the macroeconomy. We next turn to some theoretical insights that may shed light on these effects.

3. Weak Links and the Future of Economic Growth

Task-based models provide key insights to help us understand the effects of automation in general, and AI-style automation in particular.[2] In these models, production depends on the successful completion of a number of tasks. Initially, tasks are performed by labor. Automation occurs when firms figure out how to use machines in place of labor to perform particular tasks more cheaply. As Zeira (1998) emphasizes, automation has been going on at least since the Industrial Revolution of the 19th century, and actually for hundreds of years. A classic 19th-century example is the replacement of labor in weaving textiles by mechanical looms. In the 20th century, cars and trucks replace people and horses in transportation, tractors replace people in many agricultural tasks, and electronic computers replace human computers in performing calculations.

In studying the effect of automation on economic growth, Aghion, Jones, and Jones (2019) focus on the case in which the production function depends on tasks that combine with an elasticity of substitution less than one. That is, the model is one of "weak links." Each task can be thought of as a link in the chain, and all tasks are essential to

[2]Useful starting points in this literature include Autor, Levy, and Murnane (2003), Acemoglu and Restrepo (2018), Hemous and Olsen (2022), B. Jones and X. Liu (2024), and Korinek and Suh (2024).

production. Similarly, just as a chain is only as strong as its weakest link, production is constrained by the weakest links — the least productive tasks (Kremer, 1993; Jones, 2011).

An iconic illustration of weak links arose with the accident that destroyed the space shuttle Challenger in 1986. The post-crash investigation found that the key design flaw was in the O-rings: doughnut-shaped pieces of rubber that are designed to create a seal between two interlocking parts. Challenger launched during record-low temperatures causing the O-rings to become brittle and leak, which caused the space shuttle to explode. Partly inspired by this episode, Kremer (1993) developed his O-ring model of economic development.[3]

The consequences of weak links need not be so dramatic, and potential weak links seem pervasive in production. Consider the creation and production of a new state-of-the-art smartphone. The smartphone must be designed, the chips must be manufactured to exact specifications, a myriad of sophisticated parts must be coordinated to arrive in the right place at the right time, the smartphone needs to be delivered to the retail store, advertising, marketing, and customer management must be arranged, etc. When any of these tasks are hindered, the value of the enterprise gets reduced considerably.

Jones and Tonetti (2026) use the weak link framework to explain how the presence of weak links can limit the growth that emerges from artificial intelligence. A simple example illustrates how this works. Suppose that final output is produced by combining two tasks: $Y = F(X_{easy}, X_{hard})$. The "easy" task is easily automated, while the "hard" task is hard to automate. To keep things simple, let's suppose that $X_{hard} < X_{easy}$ — the hard task is also harder to produce and so constitutes the smaller input.

As in much of the task literature, assume this production function features a constant elasticity of substitution (CES). It simplifies things considerably to further assume that the elasticity of substitution is equal to 1/2. In this case, output is proportional to

[3]Kremer's O-ring model is based on a Leontief production function, i.e. a CES production function with a zero elasticity of substitution. The weak link production function allows the importance of weak links to be a parameter, as discussed in the next several paragraphs and the next footnote.

the harmonic mean:[4]

$$\frac{1}{Y} = \frac{1}{X_{easy}} + \frac{1}{X_{hard}}$$

In this formulation, the hard task is the weakest link and it limits total production. That is, if $X_{easy} = \infty$, then $Y = X_{hard}$, so with less than infinite X_{easy}, we have that $Y < X_{hard}$. Notice that this also means that even with infinite amounts of some input, overall production remains finite — again because we are limited by the weakest link. These results generalize to a production function involving many tasks. In this sense, it becomes clearer how automating many tasks might not lead to huge output gains: output is always constrained by the weakest links that are not yet automated.

To see a familiar example of this phenomenon, just look at your computer. You have on your desk roughly 100 million times the computing power that was available on the best computers from the early 1970s. But you and I are not 100 million times more productive than the economists of that time. Matrices get inverted much more rapidly, but we still have to decide what data to put in the matrix, what the structure of the economic model is, and what theory we are testing. In other words, our production is limited by the weak links.

Now return to the harmonic mean example and suppose that we choose units so that initial output is $Y_{initial} = 1$. What happens if we then fully automate the "easy" task? In fact, the cleanest result comes from supposing we become so good at the easy task that $X_{easy} = \infty$; it is no longer a bottleneck in any way. In that case, it is straightforward to show that output with infinite amounts of X_{easy} is given by a simple formula: $\frac{1}{1-s}$ where s is the original spending share on the easy task, assuming competitive markets.[5] That is, infinitely automating a task that costs the economy a fraction of GDP given by s raises output by the factor $\frac{1}{1-s}$.

Next, recall the "AI accelerates economic growth" scenario. One of the first pillars of that scenario is that AI automates software production. How large an effect would this have on the economy? Well, even with *infinite* amounts of software doing the

[4]The standard CES production function in this case would be $Y = (X_{easy}^{\frac{\sigma-1}{\sigma}} + X_{hard}^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}}$. Setting $\sigma = 1/2$ delivers the expression in the text. The standard arithmetic mean occurs if $\sigma = \infty$ so the exponents become one (we are ignoring the 1/2 multiplier that is needed to turn the sum into the mean). The geometric mean (Cobb-Douglas) occurs when $\sigma = 1$, although this is hard to see because the exponents become 0 and ∞. The harmonic mean is just another way of measuring the central tendency; it puts more weight on the smaller values — that is, on the weak links.

[5]Jones and Tonetti (2026) derive this formula for a more general constant-elasticity-of-substitution setup; see also B. Jones (2025).

existing set of tasks that software performs today, output would rise by $\frac{1}{1-s}$. Spending on software accounts for around 2 percent of GDP, and for small s, $\frac{1}{1-s} \approx 1+s$. In other words, having access to infinite amounts of the tasks that software performs today would only raise GDP by around 2 percent. Why is the gain so small? Weak links.

In the more general case where the elasticity of substitution — the weak link parameter — is σ, output after infinitely automating the share s of GDP is given by

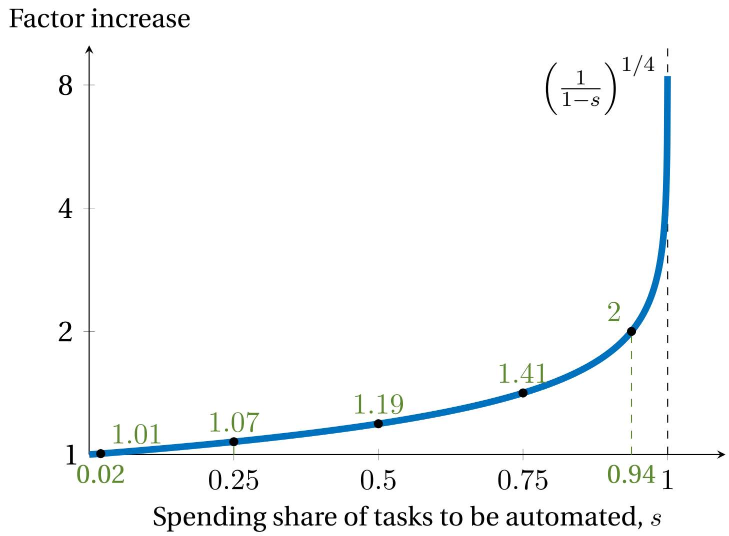
$$\left(\frac{1}{1-s}\right)^{\frac{\sigma}{1-\sigma}}$$

When $\sigma = 1/2$, this reduces to the simple case where the exponent is one. To the extent that σ is smaller, weak links are more important. For example, if $\sigma = 0.2$, then the exponent becomes $\frac{0.2}{0.8} = 1/4$ and the output gains from infinite automation will be even smaller.

Figure 2 illustrates this formula graphically. That is, the figure shows the proportional gain from infinitely automating more and more tasks for the case in which weak links are significant: $\sigma = 0.2$, which is the value chosen in Jones and Tonetti (2026) based on a range of evidence. The horizontal axis is s, the share of GDP that gets automated. There are two key messages from the figure. First, as the share of the economy that get automated increases, the gains are larger. Indeed, if one hundred percent of the economy is automated, GDP would be infinite. This is an intuitive explanation for how the gains from automation can ultimately be very large — as in the "AI accelerates growth" scenario considered earlier.

The second message from the figure is that weak links can significantly slow the gains from automation. In the "AI accelerates growth" scenario, the next step after automating software is that AI automates all cognitive work that can be performed on a computer. To see the consequences of this, consider the case in which $s = 1/2$ so that 50 percent of GDP gets infinitely automated. According to Figure 2, this would raise GDP by just 19 percent. The reason is once again weak links. With $\sigma = 0.2$, having infinite amounts of half the tasks in the economy has relatively small effects because we are limited by the remaining tasks that are performed by humans. We are still using people to educate our children, provide healthcare, run the government, manufacture goods, and provide hotels, restaurants, and travel experiences. Indeed, even just doubling income per person requires the infinite automation of 94 percent of GDP.

Figure 2: GDP per Person with Infinite Automation



Note: The figure shows the factor increase in GDP per person from automating with infinite productivity tasks that have an initial cost share of GDP equal to s, assuming an elasticity of substitution equal to 0.2 (Jones and Tonetti, 2026).

Of course, the set of tasks that can be automated in the economy is not frozen in time but instead evolves as we discover new ideas. Jones and Tonetti (2026) build a dynamic model in which the production of both goods and ideas gets progressively automated over time. More automation leads to more new ideas, which in turn leads to further automation, and new ideas are continually making machines more and more productive at the tasks they perform. On the other hand, at any point in time this flywheel dynamic is constrained by weak links.

The equation graphed in Figure 2 can be viewed as an extremely simplified version of that much richer model. In particular, suppose that we begin the economy at $s = 0$ so there is no infinite automation. And suppose that each year s increases by .01, so that after 100 years all tasks will be infinitely automated. In that case, the curve in Figure 2 can be interpreted as the time series of income per person over the next century. And because the graph is plotted on a log scale, the slope of the curve reflects the growth rate over time. The slope is relatively stable for the first 50 or 60 years — suggesting that growth does not accelerate by much. We might expect growth rates to be similar to the 2% per year growth shown in Figure 1. After 75 years, however, the slope starts to increase, and over the last 25 years the slope rises dramatically, and GDP becomes infinite after a century has passed.

What is the lesson? The positive feedback loop between automation and the discovery of new ideas does indeed accelerate growth, as in the first scenario considered earlier in this essay. However, the presence of weak links means that the acceleration is initially slow, as in the "business as usual" scenario. It is only after the vast majority of weak links have been automated away that the growth acceleration truly becomes evident. These same results emerge in the richer and fully dynamic model in Jones and Tonetti (2026). On the one hand, weak links slow the economic gains from AI and automation; the consequences could easily be modest for the next decade or two. On the other hand, if AI is ultimately able to automate nearly every task that humans perform, even the model that emphasizes substantial weak links features a huge acceleration in economic growth.

An example that I find helpful in illustrating this point is self-driving cars. The first DARPA Grand Challenge was in 2004. Fifteen self-driving vehicles competed and none finished the 142-mile course; the furthest any vehicle traveled was 7.5 miles (DARPA, 2014). Twenty-two years later in 2026, Waymo self-driving cars are easy to spot throughout San Francisco. The vehicles have travelled a total of 170.7 million miles without a human driver, and the record suggests they are much safer than most human drivers (Waymo, 2026). And yet San Francisco is the exception not the rule. In the vast majority of other places in the United States, you still cannot hail a self-driving car. This is another illustration of weak links (weather conditions, the long tail of edge cases, cost of sensor suites and computing, etc.) and the time horizon that is required for those links to be overcome. I suspect the same thing will be true about many other aspects of AI interacting with the physical world. We will need multiple "nines" of reliability before we have robots managing our kindergarten classrooms.

A range of possibilities exist between the "AI accelerates economic growth" and "AI is business as usual" scenarios, and other research has also considered these possibilities. Nordhaus (2021) asks whether an economic "singularity" of extraordinarily rapid growth is near and concludes that it is not. Acemoglu (2024) suggests that the macroeconomic impacts of AI may be quite modest over the next decade, raising growth in total factor productivity by less than 0.1 percentage points per year. Aghion and Bunel

(2024) question some of the empirical choices made by Acemoglu and calculate a larger gain, with AI perhaps raising growth of total factor productivity by 0.7 percentage points per year. Jones and Tonetti (2026) suggest that the "business as usual" scenario may be most relevant for the next decade or two but that it is likely that AI does indeed ultimately accelerate economic growth in the long run.

4. Labor Markets, Inequality, and Meaningful Work

Many of the concerns about artificial intelligence arise from its potential effects on labor markets. This important topic is further from my expertise, so I will offer only a few scattered remarks. Others such as Acemoglu and Autor (2011), Acemoglu and Restrepo (2022) and Autor and Thompson (2025) have studied this question closely.[6]

A key insight comes from appreciating that jobs are bundles of tasks. AI can therefore reduce wages in some jobs but raise wages in others, leading to nuanced and unexpected results. A classic example comes from a remark in 2016 by Geoffrey Hinton — who would later share the 2024 Nobel Prize in Physics "for foundational discoveries and inventions that enable machine learning with artificial neural networks" — that medical schools should stop training radiologists because AI would soon be much better at reading medical imaging scans, putting radiologists out of work. Nearly ten years later, we have *more* radiologists rather than fewer, and salaries for radiologists have risen rather than fallen (Mousa, 2025). It turns out that radiologists do more than just read scans, and AI tools complement those other skills by automating a fraction of the tasks that radiologists perform. In a weak links world, automation will raise the task prices of the remaining weak links, which in turn can raise wages, at least as long as those weak links are themselves not automated away. Automation of *some* of the essential tasks performed by radiologists raises the productivity of radiologists overall. Of course, some jobs — perhaps even radiologists some day — may consist of bundles of tasks that are all automated, and the wages of those jobs will fall to the marginal cost of performing those tasks with machines and software, which may indeed be low.

One of the questions at the start of this essay was "What happens if AI plus robots can perform every task a human can, at least as cheaply?" A first observation is that the

[6] See also Althoff and Reichardt (2025) and Freund and Mann (2025).

weak links framework described in the previous section suggests that this make take decades to occur rather than happening in the next few years. So there is perhaps time for the labor market to adjust, as the changes come gradually rather than suddenly.

Beyond the rate of adjustment, the total computing power of the economy at any point in time is finite, so the principle of comparative advantage comes into play. There will always be tasks that are scarce and that humans can perform; the question is simply at what price. But as the economy gets more productive, the wage for some tasks must rise (Caselli and Manning, 2019; Restrepo, 2025). Humans are not getting *worse* at performing tasks, and some tasks will remain scarce. None of this is to say that some workers cannot see their wages fall. Perhaps humans leave highly-paid occupations like software engineers and lawyers in order to take lower paying jobs that retain value precisely because they are performed by humans (e.g. artists and musicians). The wage can fall when people switch jobs even when the wage for artists is always rising. Similarly, even if in some scenarios the share of GDP paid to labor is falling to zero, this does not mean that the wage is falling. For example, if total labor income rises at 3% per year while total capital income rises at 5% per year, the labor share will fall to zero. But labor incomes are rising. All of this is simply to reinforce that the labor market consequences of AI can be nuanced.

Another point worth appreciating is that a world in which artificial intelligence combined with robotics can automate nearly all tasks is a world of abundance. The “size of the pie” could become very large, as we have already discussed, and in principle the large increases in GDP could make everyone better off: it becomes a question of distribution. Historically, the main asset that many people have is their labor endowment, and renting this asset to an employer was (and is) the way that many people earned income. Thus, an important question is what economically valuable endowment people possess that will enable them to share in any riches that artificial intelligence creates. Modern economies engage in substantial redistribution through a system of taxes and benefits, and large gains in GDP would make this kind of redistribution more affordable. Of course, the political economy of redistribution could easily become more complicated to the extent that wealth gets even more concentrated — say, if the financial benefits of discovering advanced AI flow to a small set of people. Questions around the distribution of income are likely to become very important in the future and deserve careful consideration. Trammell and Patel (2025) offer a readable starting point for discussing the potential broader consequences of AI for wealth inequality.

In most economic models, work is a “bad” rather than a “good” — which is why people must be paid to do it. However, for many people, and perhaps even for an increasing number over time, jobs provide one foundation for finding meaning in their lives. When AI is better than me at understanding the sources of economic growth and developing new growth models, where will I find meaning in my life? In early 2026, ChatGPT 5.2 Pro was already better than me at solving many of the growth problems that I work on.[7]

I don't know the answer to this question, but metaphors I've found helpful in thinking about it are *retirement* and *summer camp*. Retirees with ample incomes and plentiful time seem to find meaning in life by spending time with friends, travel, new experiences, etc. Perhaps in the future, my growth economist friends and I will gather in interesting locales and have AI teach us the latest growth models that it has discovered. Most of what we enjoy about research is learning new things, and this will only get easier as AI improves. In the end, advanced AI will perhaps understand humanity better than we understand ourselves, and we can ask it for advice about how best to live a meaningful life in a brave new world.

5. Catastrophic Risk

Modern-day Prometheans like Sam Altman of OpenAI, Dario Amodei of Anthropic, Demis Hassabis of Google DeepMind, and Geoff Hinton, who is sometimes known as the “Godfather of AI” for his work on neural networks, were all early advocates of the incredible promise of artificial intelligence. But they share another important characteristic as well. Before 2020, they all warned about the significant potential risks of AI technologies: a common message was that AI may be a larger economic driver than electricity or the internet, but could also be more dangerous than nuclear weapons. Indeed, OpenAI was founded explicitly as a nonprofit so that it could focus on developing AGI safely, avoiding pressures from market incentives to develop powerful AI models

[7] For example, see this problem of characterizing explosive economic growth: https://web.stanford.edu/~chadj/explosivegrowthChatGPT5.2.pdf.

before safety measures were fully in place.[8]

Risks from artificial intelligence broadly fit into two categories. The first is associated with *bad actors*. Consider AI models in the next five to ten years. Extrapolating from recent gains, we could have amazing AI models that could master biochemistry and virology. A bad actor could ask such a model to develop the recipe for a novel virus that is more deadly and more contagious than Ebola, but that takes four weeks for symptoms to emerge. The modern world has so far managed to avoid the worst outcomes associated with nuclear weapons in part because only a small number of parties had access to the "red button." Advanced AI could mean we have millions of people with access to red buttons.

The second category of catastrophic risk is more speculative and might be called *alien intelligence*. We are growing new forms of intelligence that we do not understand. Suppose we found out tomorrow that an alien spacecraft was passing Pluto on its way to Earth. How would we feel? It would surely be exciting, but we might also feel some trepidation: when more advanced species or societies encounter less advanced ones historically, it does not usually end well for the less advanced party. Stuart Russell, a Berkeley computer science professor who is coauthor of one of the leading graduate textbooks on artificial intelligence, has offered a thought-provoking comment: "How do we retain power over entities more powerful than us, forever?"[9]

These risks are inherently speculative. For present purposes, a natural question is this: To what extent can economic analysis shed light on these risks given the inherent uncertainties? The remainder of this section discusses some progress economists have made along these lines.

5.1 The Oppenheimer Question

In anticipation of the first test of the atomic bomb, the scientists of the Manhattan Project faced potentially catastrophic risks: for example, what if the nuclear chain reaction continued unabated, igniting the atmosphere and potentially killing all life on Earth? The scientists considered this question, as explained in a now-declassified re-

[8]For discussions of catastrophic risks from AI and other sources, useful starting points are Bostrom (2002), Rees (2003), Posner (2004), Yudkowsky et al. (2008), and Nielsen (2024). And of course OpenAI is no longer a pure nonprofit. Ord (2020), in Table 6.1, attempts to quantify other sources of existential risk as well. Unaligned artificial intelligence is by far the largest risk, followed by engineered pandemics.

[9]From a panel presentation in Anton Korinek's CEPR AI online seminar, February 25, 2025.

port by Konopinski, Marvin, and Teller (1946). They estimated the probability to be extremely low, and the Trinity test went forward. But how large would the risk have to be to avert the test?

Jones (2024) considers an analogous question with respect to artificial intelligence. Suppose on one side that AI has incredible benefits, raising economic growth to 10 percent per year and thus doubling the average standard of living every 7 years. However, it comes with a one-time risk of killing everyone on the planet. How much risk would standard economic agents be willing to take in this case?

Several surprising results emerge. First, consider the situation in which agents have log utility and begin with living standards comparable to the US average. With log preferences, the marginal utility of consumption is proportional to the inverse of consumption — a doubling of consumption leads marginal utility to fall by 50 percent. In this case, it turns out that the agent would be willing to take any risk lower than 1-in-3 of killing everyone in order to get 10 percent growth. (I learned from this calculation that I do not have log utility!)

Next, consider a situation in which marginal utility falls with the *square* of consumption, which is a case often considered in macroeconomics. Now, the marginal utility of consumption falls twice as fast. The gain from 10 percent annual consumption growth is worth much less, and the acceptable risk plummets: agents are only willing to accept a 2.5 percent existential risk in exchange for growth. At US living standards, utility is relatively high already and the marginal utility of additional consumption is relatively low. It is not worth taking large risks to losing our high living standards to get gains in consumption that are not very valuable.

The final surprise in Jones (2024) comes from recognizing that if AI is remarkable enough to deliver 10 percent economic growth, it will probably also produce tremendous advances in medicine — perhaps curing cancer and heart disease, and offering a dramatic extension of healthy lifespans. Suppose that the new AI future cuts mortality rates in half (admittedly a large change, but then, so is the shift to 10 percent growth). The agents must now include in their calculations that in the good scenario, AI will not only improve consumption but will also dramatically reduce the chance of death and extend life expectancy. In this case, it turns out that even when the marginal utility of consumption falls rapidly, agents are willing to take large existential risks.

The relevant tradeoff is not consumption versus death but rather death from cancer versus death from misaligned AI (in which case the marginal utility of consumption is no longer particularly relevant). Although agents care about not dying, they don't care about the cause of death. In this case, no matter how fast the marginal utility of consumption declines, standard economic agents are willing to take a 1-in-4 chance of killing everyone in order to cut mortality rates in half. Clearly, the health and medical benefits of artificial intelligence may be particularly valuable.

These thought experiments obviously ignore many important considerations. Perhaps we assign a value to humanity continuing over and above the selfish benefits considered in these examples. Nevertheless, the calculations may be informative, particularly about the types of AI models that are most valuable — for example, medical innovations may be especially important.

5.2 How much spending to mitigate existential risk?

Of course, just because people are *willing* to accept a certain amount of risk does not mean that we should do nothing. What if we can undertake actions to reduce the existential risk itself? How much should we be spending to mitigate existential risk (Jones, 2025)?

A point emphasized by Ord (2020) and MacAskill (2022) is that an existential risk that ends humanity would arguably prevent the existence of thousands of future generations, constituting trillions of future people. When a generation stands at the precipice of a decision that will have substantial consequences for future generations, it arguably has a moral obligation to undertake large investments to limit the existential risk. This point is clearly important, but it also involves a moral argument that we should give substantial weight to people in distant future generations who do not yet exist — and indeed, may never exist.

Our recent experience with the Covid-19 pandemic suggests a very different motivation for spending to mitigate existential risk. In particular, in 2020, we each faced an imminent mortality risk of something like 0.3 percent from Covid-19. As a society, we responded by shutting down the economy and remaining in our homes, "spending" the equivalent of around 4 percent of GDP in the United States to mitigate this risk to current generations (Goolsbee and Syverson, 2021; Fernández-Villaverde and Jones, 2020).

If one believes the catastrophic risks from artificial intelligence are at least this large, then by revealed preference perhaps we should be spending an equivalent amount, even from a purely selfish standpoint that places no value on future generations.

A counterargument is that what we did in 2020 may not have been optimal. However, a simple calculation along the lines of Hall, Jones, and Klenow (2020) suggests that this consideration strengthens rather than weakens the argument. US government agencies implementing safety policies routinely use numbers on the order of $10 million or more for the value of life for an average American today (U.S. Environmental Protection Agency, 2024; U.S. Department of Transportation, 2025). To avoid a mortality risk of 1 percent, this value implies a willingness to pay of $1 \text{ percent } \times \$10 \text{ million } = \$100,000$. Average US GDP per person is around $90,000, so if projected over the entire population, this willingness to pay to reduce mortality risk by 1 percent is more than 100 percent of GDP. If the existential risk is realized in the next 10 to 20 years, an annual investment of 5–10 percent of income could be appropriate, at least if it would completely eliminate the risk.

This willingness to pay needs to be adjusted by a measure of the effectiveness of the mitigation spending, which for artificial intelligence has considerable uncertainty. However, some forms of mitigation could be effective. For example, many of the early efforts by DeepMind were *narrow* AI models such as AlphaFold. We could focus government support of AI research on narrow models that accelerate scientific research, especially in medicine, on the grounds that narrow AI may pose smaller risks. Alternatively, we could put greater emphasis on AI safety research. Algorithms that improve AI safety can be thought of as global public goods; there are surely large spillovers both across countries and over time from these types of ideas. Jones (2025) conducts various robustness checks and shows that even with a wide range of the effectiveness of mitigation spending, we are likely underinvesting in AI safety by a factor of 30 or more. Investments on the order of $100 billion per year — a third of a percent of US GDP — easily pass cost-benefit analysis criteria.

At current US levels of consumption, life is incredibly valuable and the marginal utility of consumption is correspondingly low. From a purely selfish standpoint — placing no weight on future generations — it is worth spending perhaps surprisingly large amounts of money to mitigate risks to life based on valuations that the US government uses every day.

5.3 An AI Arms Race and Policy

It feels as if there is a prisoners' dilemma element to AI race dynamics at the moment. On the one hand, many of the leaders of the AI labs have historically warned about the potential risks associated with AI. On the other hand, these same leaders seem to be racing ahead to build data centers and advance the technology before safety problems are solved. One can imagine each lab individually saying: Everyone else is racing. If I slow down, that does not meaningfully change the existential risks. But if I race, then (a) maybe I will be safer than the others, and (b) enormous gains await the winner of the race if the catastrophic risks are not realized. The equilibrium is that everyone races, even though everyone may be better off if all slowed down.

If the main parties understand the negative aspects to arms race dynamics, it raises the possibility that international cooperation, mediated and checked by third parties, could be possible. For example, the 1970 Treaty on the Non-Proliferation of Nuclear Weapons seems to have slowed the spread of nuclear weapons in the last half-century, and the 1987 Montreal Protocol on Substances That Deplete the Ozone Layer coordinated a global reduction in emissions of chlorofluorocarbons and other substances that diminished levels of ozone in the atmosphere.

6. Concluding Thoughts

How much will A.I. change the world in the coming decades? A lesson from economic history is that general purpose technologies like the internet take multiple decades to have their full impact, and surely the same will be true of artificial intelligence. Just because the effects have been modest for the past decade does not mean that the cumulative effect over half a century will be small.

The weak link framework helps us to understand this statement and to reconcile the two scenarios at the beginning of this essay. There is indeed a positive feedback loop between innovation and automation. Advances in AI help us create new algorithms and new ideas, and these new ideas in turn help us advance AI and automation. This positive feedback loop provides an impetus for economic growth to accelerate. But weak links constrain the acceleration. We are not 100 million times more productive simply because we use today's computer chips instead of chips from the 1970s. Many other necessary tasks for production have not seen this kind of stunning improvement and therefore serve as weak links that constrain the economy's performance. But if AI and AI-run robots manage to automate nearly all of the economy, the positive feedback loop could indeed accelerate economic growth to unprecedented rates, possibly 5 percent per year or even higher.

The weak link view suggests that the substantial benefits from AI might be delayed by decades as a large fraction of the weak links must be automated away. But the weak link view has the opposite implication for catastrophic risks. When a chain is only as strong as its weakest link, cutting one of the links is sufficient to do tremendous damage to the value of the chain. AI automation of software engineering in the near future may not lead to large benefits in the short term because of the other weak links. But powerful AI could do enormous damage to the economy by hacking the electricity grid or the financial system or helping a bad actor to engineer a deadly pandemic. Because of weak links, the benefits from AI may arrive gradually but some risks may arrive quickly. It would be prudent to spend the intervening time making preparations for the potentially large consequences for labor markets, inequality, and catastrophic risk.

References

Acemoglu, Daron, "The Simple Macroeconomics of AI," Working Paper 32487, National Bureau of Economic Research May 2024.

_ and David Autor, "Skills, Tasks and Technologies: Implications for Employment and Earnings," in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Vol. 4, Elsevier, June 2011, chapter 12, pp. 1043–1171.

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

_ and _ , "Tasks, Automation, and the Rise in U.S. Wage Inequality," *Econometrica*, 2022, *90* (5), 1973–2016.

Aghion, Philippe and Simon Bunel, "AI and Growth: Where Do We Stand?," June 2024. unpublished manuscript.

_ , Benjamin F. Jones, and Charles I. Jones, "Artificial Intelligence and Economic Growth," in Ajay Agrawal, Joshua Gans, and Avi Goldfarb, eds., *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, 2019, pp. 237–282.

Althoff, Lukas and Hugo Reichardt, "Task-Specific Technical Change and Comparative Advantage," 2025. Unpublished manuscript.

Amodei, Dario, "Machines of Loving Grace: How AI Could Transform the World for the Better," October 2024.

Anthropic, "Introducing Claude Opus 4.5," November 2025. https://www.anthropic.com/news/claude-opus-4-5.

Aschenbrenner, Leopold, "Situational Awareness: The Decade Ahead," June 2024.

Autor, David and Neil Thompson, "Expertise," *Journal of the European Economic Association*, 06 2025, *23* (4), 1203–1271.

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

Barro, Robert J. and José F. Ursua, "Barro-Ursua Macroeconomic Data," 2010. Downloaded from https://scholar.harvard.edu/barro/publications/barro-ursua-macroeconomic-data.

Bloom, Nicholas, Charles I. Jones, John Van Reenen, and Michael Webb, "Are Ideas Getting Harder to Find?," *American Economic Review*, April 2020, *110* (4), 1104–44.

Bostrom, Nick, "Existential Risks," *Journal of Evolution and Technology*, March 2002, *9*.

Brynjolfsson, Erik and Lorin M Hitt, "Beyond computation: Information technology, organizational transformation and business performance," *Journal of Economic perspectives*, 2000, *14* (4), 23–48.

_ , Anton Korinek, and Ajay K. Agrawal, "A Research Agenda for the Economics of Transformative AI," Working Paper 34256, National Bureau of Economic Research September 2025.

Caselli, Francesco and Alan Manning, "Robot Arithmetic: New Technology and Wages," *American Economic Review: Insights*, June 2019, *1* (1), 1–12.

Cunningham, Tom, "Economics and Transformative AI," November 2025. https://tecunningham.github.io/posts/2025-09-19-transformative-AI-notes.html.

DARPA, "The DARPA Grand Challenge: Ten Years Later," https://www.darpa.mil/news/2014/grand-challenge-ten-years-later March 2014. Defense Advanced Research Projects Agency, Tactical Technology Office. Accessed 2026-05-15.

David, Paul A., "The Dynamo and the Computer: An Historical Perspective on the Modern Productivity Paradox," *American Economic Association Papers and Proceedings*, May 1990, *80* (2), 355–361.

Davidson, Tom, "Could Advanced AI Drive Explosive Economic Growth?," June 2021. Open Philanthropy report.

—, Basil Halperin, Thomas Houlden, and Anton Korinek, "Is Automating AI Research Enough for a Growth Explosion?," 2026. unpublished manuscript.

Emberson, Luke, "Frontier AI Capabilities Accelerated in 2024," Epoch AI Substack December 2025. Accessed: 2026-05-18.

Epoch AI, "Data on AI Models," 07 2025. Accessed: 2025-12-18.

—, "GATE — AI and Automation Scenario Explorer," 2025.

Fernández-Villaverde, Jesús and Charles I. Jones, "Macroeconomic Outcomes and COVID-19: A Progress Report," *Brookings Papers on Economic Activity*, Fall 2020, pp. 111–166.

Freund, Lukas B. and Lukas F. Mann, "Job Transformation, Specialization, and the Labor Market Effects of A," 2025. unpublished manuscript.

Goolsbee, Austan and Chad Syverson, "Fear, lockdown, and diversion: Comparing drivers of pandemic economic decline 2020," *Journal of Public Economics*, 2021, *193*, 104311.

Hall, Robert E., Charles I. Jones, and Peter J. Klenow, "Trading Off Consumption and Covid-19 Deaths," *Quarterly Review*, 2020, *42* (1).

Hemous, David and Morten Olsen, "The Rise of the Machines: Automation, Horizontal Innovation, and Income Inequality," *American Economic Journal: Macroeconomics*, January 2022, *14* (1), 179–223.

Ho, Anson, Tamay Besiroglu, Ege Erdil, David Owen, Robi Rahman, Zifan Carl Guo, David Atkinson, Neil Thompson, and Jaime Sevilla, "Algorithmic progress in language models," 2024.

Jones, Benjamin, "Artificial Intelligence in Research and Development," NBER Working Paper 34312 October 2025.

__ and Xiaojie Liu, "A Framework for Economic Growth with Capital-Embodied Technical Change," *American Economic Review*, May 2024, *114* (5), 1448–87.

Jones, Charles I., "Intermediate Goods and Weak Links in the Theory of Economic Development," *American Economic Journal: Macroeconomics*, April 2011, *3* (2), 1–28.

__ , "The AI Dilemma: Growth versus Existential Risk," *American Economic Review: Insights*, December 2024, *6* (4), 575–90.

__ , "How much should we spend to reduce A.I.'s existential risk?," March 2025. NBER Working Paper 33602.

__ and Christopher Tonetti, "Past Automation and Future A.I.: How Weak Links Tame the Growth Explosion," January 2026. unpublished manuscript.

Kokotajlo, Daniel, Scott Alexander, Thomas Larsen, Eli Lifland, and Romeo Dean, "AI 2027," 2025. https://ai-2027.com/.

Konopinski, Emil J., Cloyd Marvin, and Edward Teller, "Ignition of the Atmosphere with Nuclear Bombs," Technical Report LA-602, Los Alamos National Laboratory August 1946. August 14, 1946.

Korinek, Anton and Donghyun Suh, "Scenarios for the Transition to AGI," Working Paper 32255, National Bureau of Economic Research March 2024.

Kremer, Michael, "The O-Ring Theory of Economic Development," *Quarterly Journal of Economics*, August 1993, *108* (4), 551–576.

MacAskill, William, *What We Owe the Future*, Basic books, 2022.

METR, "Task-Completion Time Horizons of Frontier AI Models," https://metr.org/time-horizons/ May 2026.

Moore, Gordon E., "Cramming More Components onto Integrated Circuits," *Electronics Magazine*, April 1965, *38* (8).

Mousa, Deena, "The algorithm will see you now," *Works in Progress*, September 2025.

Narayanan, Arvind and Sayash Kapoor, "AI as Normal Technology," *Knight First Amendment Institute*, April 2025.

Nielsen, Michael, "How to be a wise optimist about science and technology?," December 2024. https://michaelnotebook.com/optimism/index.html.

Nordhaus, William D., "Are We Approaching an Economic Singularity? Information Technology and the Future of Economic Growth," *American Economic Journal: Macroeconomics*, January 2021, *13* (1), 299–332.

Ord, Toby, *The Precipice: Existential Risk and the Future of Humanity*, Hachette Books, 2020.

Posner, Richard A., *Catastrophe: Risk and Response*, Oxford University Press, 2004.

Rees, Martin, *Our Final Century*, London: William Heinemann, 2003.

Restrepo, Pascual, "We Won't be Missed: Work and Growth in the AGI World," October 2025. Prepared for the NBER conference volume on The Economics of Transformative AI.

Solow, Robert M., "We'd better watch out," *New York Times Book Review*, July 1987, *36*.

Trammell, Philip and Dwarkesh Patel, "Capital in the 22nd Century," December 2025. https://substack.com/inbox/post/182789127.

U.S. Bureau of Economic Analysis, "National Income and Product Accounts," 2025. https://www.bea.gov/products/national-income-and-product-accounts.

U.S. Department of Transportation, "Departmental Guidance on Valuation of a Statistical Life in Economic Analysis," Technical Report 2025. Accessed January 10, 2025.

U.S. Environmental Protection Agency, "Mortality Risk Valuation," Technical Report 2024. Accessed December 30, 2024.

Waymo, "Safety Impact," https://waymo.com/safety/impact/ 2026. Data through December 2025. Accessed 2026-05-15.

Yudkowsky, Eliezer et al., "Artificial intelligence as a positive and negative factor in global risk," *Global catastrophic risks*, 2008, *1* (303), 184.

Zeira, Joseph, "Workers, Machines, And Economic Growth," *Quarterly Journal of Economics*, November 1998, *113* (4), 1091–1117.