A.I. and Our Economic Future

Charles I. Jones*

Stanford GSB and NBER

January 15, 2026 — Version 1.0

Abstract

Artificial intelligence (A.I.) will likely be the most important technology we have ever developed. Technologies such as electricity, semiconductors, and the internet have been transformative, reshaping economic activity and dramatically increasing living standards throughout the world. In some sense, artificial intelligence is simply the latest of these general purpose technologies and at a minimum should continue the economic transformation that has been ongoing for the past century. However, the case can certainly be made that this time is different. Automating intelligence itself arguably has broader effects than electricity or semiconductors. What if machines — A.I. for cognitive tasks and A.I. plus advanced robots for physical tasks — can perform every task a human can do but more cheaply? What does economics have to say about this possibility, and what might our economic future look like?

*In preparation for the *Journal of Economic Perspectives*. I'm grateful to Ben Jones, Tom Houlden, Chris Tonetti and Phil Trammell for helpful feedback and discussions.

1. Introduction

Artificial intelligence (A.I.) will likely be the most important technology we have ever developed. Technologies such as electricity, semiconductors, and the internet have been transformative, reshaping economic activity and dramatically increasing living standards throughout the world. In some sense, artificial intelligence is simply the latest of these general purpose technologies and at a minimum should continue the economic transformation that has been ongoing for the past century.

However, the case can certainly 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 — arguably has broader effects than electricity or semiconductors. 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 — A.I. for cognitive tasks and A.I. plus advanced robots for physical tasks — can perform every task a human can do but more cheaply? While this is by no means guaranteed to happen in the next several decades, such a radical transformation cannot be ruled out and indeed many observers expect that it will occur (Amodei, 2024; Aschenbrenner, 2024; Brynjolfsson et al., 2025; Kokotajlo et al., 2025).

The point of this essay is to entertain the possibility that A.I. might be profoundly transformative. What does economics have to say about this possibility, and what might our economic future look like?

We begin by outlining two extreme scenarios for the impact of A.I. on the economy: one in which A.I. drastically accelerates economic growth and another in which A.I. is "business as usual." Both scenarios are plausible, and the future will presumably lie somewhere between these extremes. Next, we describe task-based models of economic growth in which tasks are complementary. This "weak links" framework is helpful for thinking about the consequences of A.I. Finally, we consider potential risks from A.I., including effects on labor markets, inequality, and, more speculatively, existential risk.

2. Two Extremes for the Future of Growth

To begin, let me sketch out two possible futures, one in which A.I. has huge economic impacts and the other in which A.I. is more "business as usual." Both of the scenarios

I outline are possible and even plausible outcomes in my view. Each scenario will describe changes that could occur over 25 or 30 years. So the time period I have in mind is roughly one generation: what will the world look like as a child born today grows into an adult?

2.1 A.I. Accelerates Economic Growth

What would the next several decades be like if A.I. accelerates economic growth and has truly profound macroeconomic consequences? Scenarios along these lines have been presented by various authors.[1] They typically begin with A.I. raising the productivity of software engineers. By many accounts this is already solidly under way. For example, when Anthropic introduced Claude Opus 4.5, they highlighted its exceptional performance on a two-hour take-home exam they give to prospective software engineering hires: the A.I. model scored higher than any human candidate ever (Anthropic, 2025). That same model is measured to complete software engineering tasks that take humans nearly 5 hours with a 50% success rate (METR, 2025).

Epoch AI estimates that the amount of "effective compute" (e.g. total computing power adjusted for the quality of algorithms and software) used to train A.I. 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, the METR time horizon for 50% success on software engineering tasks just cited is doubling roughly every 5 to 7 months. What is 5 hours today was just 19 minutes a year and a half ago.

One of the frontier uses of A.I. models today is creating A.I. agents, models adept at doing tasks that humans normally do with computers. Beyond coding, this can also include writing and editing documents, using the internet, building spreadsheet models, developing slide presentations, and brainstorming ideas.

A.I. models raise the productivity of software engineers and A.I. researchers, allowing them to develop even better models in the future. At some point, it is possible that a future A.I. model might itself be able to engage in A.I. research — a situation where new A.I. models create even better A.I. models, known as *recursive self-improvement*.

[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).

Combining all of these forces, it is possible that sometime in the next decade or two — 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). We could potentially run billions of instances of advanced A.I. models that can accomplish nearly any of the tasks that talented humans can currently accomplish on computers.

These geniuses-in-a-data-center could be very helpful in raising the productivity of R&D more generally. We've already seen A.I. discover new ideas: the best example is AlphaFold, the A.I. 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 a Nobel Prize in chemistry for Demis Hassabis and John Jumper of Google DeepMind. A country of A.I. geniuses could design new pharmaceuticals and predict which drugs are likely to do best in clinical trials with minimal side effects. They could advise scientists on what lab experiments to run in fields ranging from biochemistry to materials science to nuclear fusion and energy. Cheap clean energy could help solve climate change problems and transform developing and rich countries alike.

Genius-level A.I. models could use virtual simulations and human-run experiments to design and build better robots, iterating until exceptional performance is realized. Eventually, this flywheel could result in A.I. models that can perform nearly all cognitive tasks and A.I.-run robots 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 be tremendous. Accelerating economic growth is certainly possible in theoretical models if A.I. continues to automate an ever increasing share of tasks (Aghion, Jones, and Jones, 2019; Jones and Tonetti, 2026).

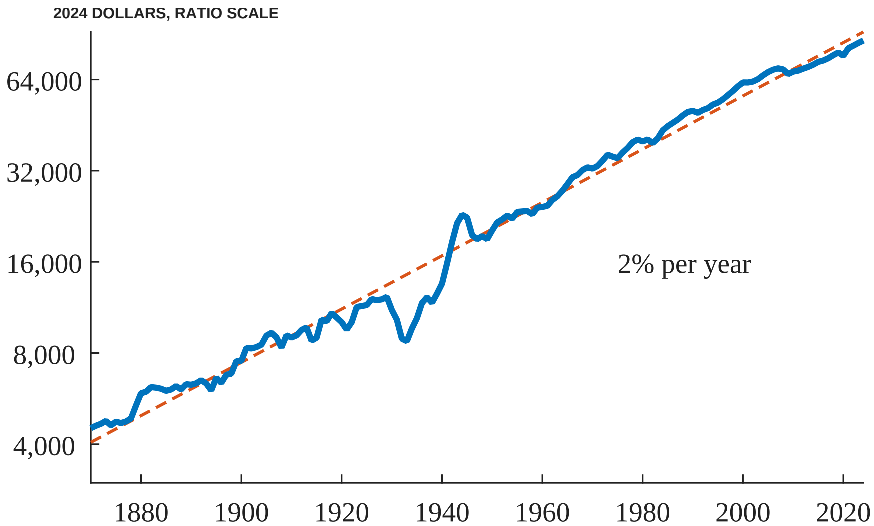
2.2 A.I. as "Business as Usual"

We've just laid out the case for A.I. as a technology that profoundly changes the economy. What about the case in which A.I. is instead "business as usual"?

Consider my favorite graph in economics, shown in Figure 1. This plot shows real GDP per person in the United States for the past 150 years. On a ratio scale, the time series is roughly a straight line — the Moore's Law of macroeconomics.[2]

[2] Recall that Moore's Law is the empirical regularity that the density of transistors on a computer chip

Figure 1: Average U.S. 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.

But also 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. Importantly, many of these are what economic historians call "general purpose technologies" (GPTs) whose transformative effects extend throughout the economy. Indeed many could also be said to automate some of the tasks involved in creating new ideas, raising the productivity of the idea production function.

And yet: apparently none of these innovations changed the long-run growth rate of the U.S. economy. How can we understand this disconnect? A 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 new general purpose technologies, one might expect economic growth to slow down.

is doubling every two years (Wikipedia, 2026). Because of the Rule of 70, Moore's Law and U.S. GDP per capita growth are in some sense duals: transistor density doubles every two years and grows at 35% per year while per capita GDP grows at 2% per year and doubles every 35 years.

From this perspective, each of these new GPTs did indeed raise the growth rate of the economy: without the next GPT, the counterfactual is that growth would have slowed considerably. The continued development of these amazing new technologies is what made sustained growth at 2% per year possible. And perhaps A.I. is just the latest GPT that lets 2% growth continue for another 50 years. Notice that even in this scenario, A.I. potentiallyhas large effects since the counterfactual is one in which growth would have slowed.

Gradual diffusion. Another complementary lesson from economic history is that it can take many decades before the impact of a new innovation shows up in productivity. David (1990) first made this point in the context of the steam engine and the electric motor. Bob Solow famously quipped in 1987 that "You can see the computer age everywhere but in the productivity statistics" (Solow, 1987). New technologies require many complementary innovations before they can have their full effect. Factories needed to be redesigned to take advantage of 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 A.I., the clear lesson from economic history is that the effects on GDP and productivity may take longer to appear than we expect.

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 A.I. And indeed, models of economic growth in which A.I. automates most of the tasks in the economy can formally produce explosive economic growth (Aghion, Jones, and Jones, 2019). So this scenario does seem possible.

On the other hand, automation 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% per year.

At some level, one of the most important things to realize is that there is substantial uncertainty about the future effects of A.I. on the macroeconomy. In either scenario, the effects are large and profound, just as they were with electricity and the internet.

Still exactly how large and how profound is far from clear. Economic theory may shed light on these questions.

3. Weak Links

Task-based models of economic growth provide key insights to help us understand the effects of automation (Zeira, 1998; Acemoglu and Restrepo, 2018).[3] In these models, production depends on the successful completion of a number of tasks. Initially, tasks are performed by labor. Automation is the process by which we figure out how to use machines / capital in place of labor to perform a particular task. As Zeira (1998) emphasizes, automation has been going on for hundreds of years, at least since the Industrial Revolution. A classic example is the replacement of labor in weaving textiles by mechanical looms. Similarly, 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.

Aghion, Jones, and Jones (2019) study the case in which tasks combine with an elasticity of substitution less than one. That is, the model is more akin to the Kremer (1993) O-ring setup than to the standard love-of-variety approach that commonly features in growth and trade models. All tasks are essential, and production is constrained by the bottlenecks or weakest links (Jones, 2011).

A simple example from Jones and Tonetti (2026) is helpful in conveying the kind of results that can emerge. Suppose final output is produced by combining the output of two tasks: $Y = F(Y_{easy}, Y_{hard})$. The "easy" task is easily automated while the "hard" task is hard to automate. To keep things simple, let's suppose that $Y_{hard} < Y_{easy}$ — the hard task is also harder to produce and so constitutes the smaller input.

Next, assume this production function is CES with elasticity of substitution equal to 1/2. In this case, output is proportional to the harmonic mean:

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

There are several key features of the weak link framework that are easy to see in this example. First, even if a task is infinitely supplied, total output Y remains finite.

[3] See also Hemous and Olsen (2022), B. Jones and X. Liu (2024), and Korinek and Suh (2024).

Second, total output is limited by the weakest link. That is, Y is no larger than the *smaller* of Y_{easy} and Y_{hard}. This is easy to see based on the first result: if $Y_{easy} = \infty$, then $Y = Y_{hard}$, so with less than infinite Y_{easy}, we have that $Y < Y_{hard}$. The hard task is the weakest link and total production is limited by the production of 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.

This last point can be made more precise. What happens if we fully automate the "easy" task? In fact, the cleanest result comes from supposing we become so good at the easy task that $Y_{easy} = \infty$; it is no longer a bottleneck in any way. In this case, it is easy to show that the proportional gain in production from having infinite amounts of Y_{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. That is, infinitely automating a task that costs the economy a fraction of GDP given by s raises output by the factor $\frac{1}{1-s}$.[4]

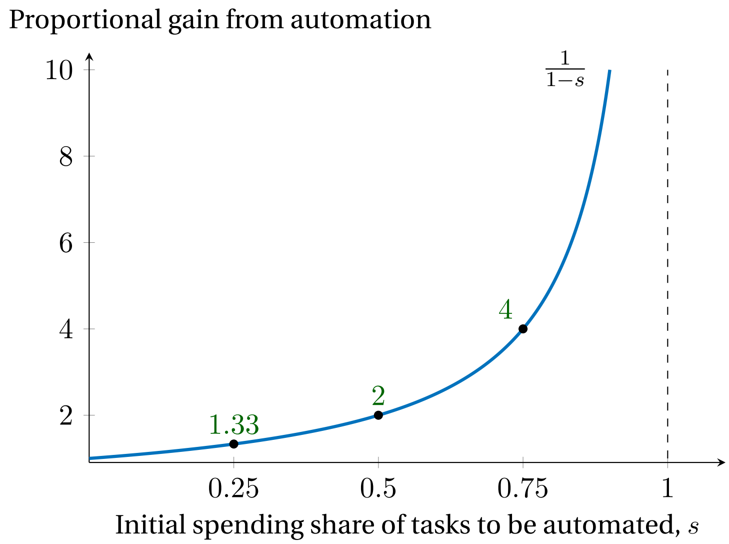
Now return to the "A.I. accelerates economic growth" scenario, and recall that one of the first pillars of that scenario is that A.I. automates software production. How large an effect would this have on the economy? Well, even with *infinite* amounts of software doing the existing set of tasks that software performs today, output would rise by $\frac{1}{1-s}$. Spending on software accounts for around 2 percent of GDP, and for s small like this, $\frac{1}{1-s} \approx 1+s$. In other words, having access to infinite amounts of the tasks that software performs today would only raise GDP by around 2 percent.

Figure 2 shows the proportional gain for tasks that occupy different initial shares of GDP. For example, knowledge work in the U.S. economy might get paid something like 1/3 of GDP. What if we automated all cognitive labor with infinite output on the tasks that it performs? This would raise GDP by 50 percent. On the one hand, if this occurred over the course of a decade, it would raise growth rates by something like 5 percent per year, which would be huge. But still, that would be a one-time gain and it is perhaps surprising that having access to infinite output of the tasks currently performed by cognitive labor might only raise GDP by 50 percent.

There are of course many qualifications to recognize in these sorts of calculations. First, this is true for an elasticity of substitution of one half. If the elasticity is smaller,

[4] Jones and Tonetti (2026) derive this formula for the more general CES setup; see also B. Jones (2025).

Figure 2: Proportional Gain from Infinite Automation



Note: The figure shows the proportional gain in GDP from automating with infinite productivity tasks that initially costs a share of GDP equal to s (Jones and Tonetti, 2026).

then the effects are even smaller; if the elasticity is larger, the effects can be much larger. For example, with Cobb-Douglas production, having infinite amounts of even a single task is enough to drive total output to infinity.

More importantly, the set of tasks being automated is not frozen in time. If software or cognitive labor were automated, one has the sense that software and cognitive labor would be used more broadly throughout the economy. Similarly, the income share s is itself an endogenous variable.

A range of possibilities exist between the "A.I. accelerates economic growth" and "A.I. is business as usual" scenarios. Nordhaus (2021) asks whether an economic "singularity" is near and concludes that it is not. Acemoglu (2024) suggests that the macroeconomic impacts of A.I. may be quite modest over the next decade, raising TFP growth by less than 0.1pp per year. Aghion and Bunel (2024) respond by questioning some of the empirical choices made by Acemoglu and calculate a larger gain, with A.I. perhaps raising TFP growth by 0.7pp per year. Jones and Tonetti (2026) construct a rich model based on the "weak link" logic to explore the future consequences of A.I. They report two findings that at first seem contradictory but are not. On the one hand, continued automation leads growth rates to rise by ever increasing amounts over the next half century, with annual growth rates eventually exceeding 5%. Nevertheless, the explosion is remarkably gradual. In 20 years, output is only higher by around 5 percent, and in forty years, it is higher by 20 percent. The insight is that weak links tame the growth explosion, making it a surprisingly gradual phenomenon.

4. Labor Markets, Inequality, and Meaningful Work

Many of the concerns about A.I. arise from its 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 Restrepo (2022) and Autor and Thompson (2025) have studied this question closely.[5] One key insight comes from appreciating that jobs are bundles of tasks. A.I. can therefore reduce wages in some jobs but raise wages in others. A classic example comes from Geoff Hinton's remark in 2016 that we should stop training radiologists because A.I. would soon be much better at reading 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). Radiologists do more than just read scans, and A.I. complements those other skills by automating a fraction of the tasks that radiologists perform. In a weak links world, automation will raise the factor income share of labor on the remaining weak links, and this can keep wages high as long as those weak links are not automated. 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. This example shows how the labor market effects of A.I. can be nuanced.

A second point worth appreciating is that a world in which A.I. automates nearly all tasks is a world of abundance. The "size of the pie" gets very large as a result of automation, 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 was (and is) the way that people

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

earned income. This will likely change in the future for many people, and an important question is what economically valuable endowment will people possess in order to share in the riches that A.I. creates. Advanced economies all engage in substantial redistribution through the tax system, and such redistribution seems likely to become even more important in the future. But at least the large gains in GDP should make this kind of redistribution more affordable: endowing every child with a share of the S&P 500 stock market index is a type of policy worth considering. Questions around the distribution of income are likely to be very important in the future and deserve further consideration by economists and others. Trammell and Patel (2025) discuss the broader consequences of A.I. for wealth inequality and is a good place to start.

In most economic models, work is a "bad" rather than a "good." That is why people must be paid to do it. However, for many people — and perhaps an increasing number over time — jobs provide a foundation for meaning. Certainly academics are very familiar with "meaningful work." When A.I. is better at understanding the sources of economic growth and developing new growth models than I am, where will I find meaning in my life? Indeed, ChatGPT 5.2 Pro is already better than I am at solving many of the growth problems that I work on.[6]

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 the A.I. 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 A.I. gets better. In the end, advanced A.I. 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 Sam Altman, Dario Amodei, Demis Hassabis, and Geoff Hinton were early advocates of the incredible promise of artificial intelligence. But they

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

all share another important characteristic in that they all warned before 2020 about the significant potential risks that A.I. could bring: A.I. may be more important 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 A.I. models before safety measures were fully in place.[7]

Risks from A.I. broadly fit into two categories. The first is associated with *bad actors*. Consider A.I. models in the next five to ten years. Extrapolating from recent gains, we could have amazing A.I. 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. Such a bioweapon could do enormous damage. 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 A.I. models could mean we have billions 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? On the one hand, this would surely be exciting. But upon reflection, 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 a quote that I find thought provoking: "How do we retain power over entities more powerful than us, forever?"[8]

Obviously these risks are inherently speculative, and 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 we have made.

[7] For discussions of catastrophic risks from A.I. and other sources, see Bostrom (2002), Rees (2003), Posner (2004), Yudkowsky et al. (2008), and Nielsen (2024).

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

5.1 The Oppenheimer Question

In anticipation of the first test of the atomic bomb, the scientists of the Manhattan Project considered potential risks: what if the nuclear chain reaction continued unabated, igniting the atmosphere and potentially killing most of life on Earth? Hans Bethe estimated the probability to be very low, and the Trinity test went forward (Wikipedia, 2025). But how large would the risk have to be to avert the test?

Jones (2024) considers an analogous question with respect to A.I. Suppose A.I. has incredible benefits, raising economic growth to 10% per year. 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, if agents have log utility, then they are willing to take any risk lower than 1-in-3 of killing everyone in order to get 10% growth. The marginal utility of consumption falls sufficiently slowly with log utility that 10% growth is extremely valuable. Next, if relative risk aversion is two instead of one (log utility), then the acceptable risk plummets: agents are now only willing to accept a 2.5% existential risk in exchange for growth. Recall that with risk aversion bigger than one, utility itself is bounded; marginal utility falls very rapidly when risk aversion is two. At U.S. living standards, life is already very valuable, and it is not worth taking large risks because the marginal utility of additional consumption is relatively low. I have a great life at Stanford: more years of life at my current standard of living is much more valuable than more consumption today.

The final surprise in the paper comes from recognizing that if A.I. is remarkable enough to deliver 10% economic growth, it will probably also give us many medical innovations — curing cancer, heart disease, and extending life. Suppose it cuts mortality rates in half (admittedly a large change but then so is the shift to 10% growth). In this case, it turns out that even very risk averse agents are willing to take large existential risks. The point is that all we care about is not dying; we don't care about what kills us. In the baseline calculation in the paper, representative agents with risk aversion of three are willing to take a 1-in-4 chance of killing everyone in order to cut mortality rates in half. One of the key lessons of the paper is that the health and medical benefits of future A.I. may be particularly valuable.

5.2 How much should we spend to mitigate existential risk?

Of course, just because we 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. Our generation stands at the precipice of a pivotal decision, and we may have a moral obligation to undertake large investments to prevent the existential risk. This is an important point, but it relies on a moral argument that we should give substantial weight to future generations who do not yet exist.

It occurred to me that our recent experience with the Covid-19 pandemic suggested a very different motivation for spending to mitigate existential risk. In particular, in 2020, we each faced a mortality risk of something like 0.3% from Covid-19. As a society, we responded by shutting down the economy and remaining in our homes, "spending" the equivalent of around 4% 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 A.I. are at least this large, by revealed preference then perhaps we should be spending an equivalent amount, even from a purely selfish standpoint.

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. U.S. 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%, this value implies a willingness to pay of $1\% \times \$10$ million $= \$100,000$. Average GDP per person is around $90,000, so this willingness to pay is more than 100% of GDP. If the existential risk is realized once in the next 10 to 20 years, an annual investment of 5–10% of income could be appropriate if it would completely eliminate the risk.

This willingness to pay needs to be multiplied by a measure of the effectiveness of the mitigation spending, and this is something there is much less certainty around. However, it is clear that some forms of mitigation could be effective. For example, many of the early efforts by DeepMind were *narrow* A.I. models such as AlphaFold. We could focus our efforts on narrow models that accelerate scientific research, especially in medicine, and this narrow A.I. may pose smaller risks that are more readily mitigated. Alternatively, we could slow down our efforts and give A.I. safety research more time to mitigate the risks. Algorithms that improve A.I. 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 A.I. safety spending by a factor of 30 or more.

The key point, again, is that at current 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 surprisingly large amounts of money to mitigate risks to life based on valuations that the U.S. government uses every day.

5.3 Race dynamics and policy

It feels as if there is a prisoners' dilemma element to A.I. race dynamics at the moment. On the one hand, many of the leaders of the A.I. labs have historically warned about the potential risks associated with A.I. 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.

One type of policy that may be worth considering is a large tax on GPUs and TPUs, the computer chips used in A.I. In addition to slowing the race, this revenue could be used to fund safety research. The tax could apply to the first sale of the chip, thereby taxing users regardless of the country in which they work.

One argument against a policy like this is that China is also racing to build advanced

A.I. and we may not want China to win the race. To the extent that China uses Nvidia chips, a chip tax could also slow down China. However, one could argue that this would incentivize Chinese chip makers who would not have to pay the tax. This would be an argument for international cooperation and a global chip tax. China and Europe also understand the negative aspects to the race dynamics, so perhaps such international cooperation, mediated and checked by third parties, would be possible, just as it has been with respect to nuclear weapons and the Cold War.

6. Concluding Thoughts

How much did the internet change the world between 1990 and 2020? How much will A.I. change the world between 2015 and 2045? 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 are modest for the first decade does not mean that the overall cumulative effect over half a century will be small. Ultimately, I expect that the effect of A.I. will be much larger than the internet, perhaps by more than 10x the internet, albeit over a half century or more. 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 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.

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.

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

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.

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/.

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, "Measuring AI Ability to Complete Long Tasks," December 2025. https://metr.org/blog/2025-03-19-measuring-ai-ability-to-complete-long-tasks/.

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

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.

Wikipedia, "Trinity (nuclear test)," December 2025. https://en.wikipedia.org/wiki/Trinity_(nuclear_test).

—, "Moore's Law," 2026. https://en.wikipedia.org/wiki/Moore%27s_law.

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.