

SCENARIO PLANNING FOR AN A(G)I FUTURE

Anton Korinek

AI may be on a trajectory to surpass human intelligence; we should be prepared



Artificial intelligence is rapidly advancing, and the pace of progress has accelerated in recent years. ChatGPT, released in November 2022, surprised users by generating human-quality text and code, seamlessly translating languages, writing creative content, and answering questions in an informative way, all at a level previously unseen.

Yet in the background, the foundation models that underlie generative AI have been advancing rapidly for more than a decade. The amount of computational resources (or, in short, “compute”) used to train the most cutting-edge AI systems has doubled every six months over the past decade. What today’s leading generative AI models can do was unthinkable just a few years ago: they can deliver significant productivity gains for the world’s premier consultants, for programmers, and even for economists (Korinek 2023).

Conjecture about AI acceleration

Recent advances in artificial intelligence have prompted leading researchers to project that the pace of current progress may not only be sustained but may even accelerate in coming years. In May 2023, Geoffrey Hinton, a computer scientist who laid the theoretical foundations of deep learning, described a significant shift in his perspective: “I have suddenly switched my views on whether these things are going to be more intelligent than us.” He conjectured that artificial general intelligence (AGI)—AI that possesses the ability to understand, learn, and perform any intellectual task a human being can perform—may be realized within a span of 5 to 20 years.

Some AI researchers are skeptical. These divergent perspectives reflect tremendous uncertainty about the speed of future progress, whether progress is accelerating or may eventually plateau. In addition, we face significant uncertainty about the broader economic implications of advances in AI and the prospective ratio of benefit to harm from increasingly sophisticated AI applications.

At a fundamental level, the uncertainty also relates to profound questions about the nature of intelligence and the capabilities of the human brain. Chart 1 shows two competing perspectives on the complexity distribution of work tasks the human brain can perform.

Panel 1 illustrates one perspective, that the capabilities of the human brain in solving ever more complex tasks are unbounded. This aligns

with our economic experience since the Industrial Revolution: as the frontier of automation advances, humans have automated simple tasks (both mechanical and cognitive) and reallocated workers to perform more of the remaining more complex tasks—that is, they have moved into the right tail of the complexity distribution illustrated in the chart. Straightforward extrapolation would suggest that this process will continue as AI advances and automates a growing number of cognitive tasks.

Another perspective, illustrated in panel 2 of Chart 1, holds that there is an upper bound to the complexity of tasks the human brain can perform. Information theory suggests that the human brain is a computational entity, constantly processing a plethora of data. The brain’s inputs include sensory perceptions—sights, sounds, and tactile sensations, among others—and its outputs manifest as physical actions, thoughts, and emotional responses. Even complex facets that make us human, such as emotions, creativity, and intuition, can be viewed as computational outputs, emerging from intricate interactions of neural circuits and biochemical reactions. Although these processes are highly elaborate and involve complexities we do not fully understand, this perspective suggests that there is a definitive upper limit to the intricacy of tasks the human brain can perform.

The two perspectives have dramatically different implications for the potential scope of future automation. As of 2023, the human brain is the most advanced computing device when it comes to the ability to perform a broad range of intellectual tasks in a robust manner. However, if the second perspective turns out to be correct, modern AI systems are catching up fast. In fact, many measures of the computational complexity of cutting-edge foundation models are already close to those of the human brain. The computational complexity of human brains is bounded by biology, and the brain’s ability to transmit information to other intelligent entities (humans or AI) is limited by the slow speed of information transmission of our senses and our language. Nevertheless, AI systems continue to advance rapidly and can exchange information at speeds that are significantly faster.

Preparing for multiple scenarios

Economists have long observed that the optimal way of dealing with uncertainty is to use a portfolio approach. Given the starkly differing perspectives on future progress in AI by world-renowned experts, it would be unwise to put all eggs in one

Robot waiters carry food for customers at a robot-themed restaurant in Chennai, India.

basket and formulate economic plans for a single scenario. Instead, the uncertainty about what the future will look like should motivate us to hedge our bets and engage in careful analysis of a range of different scenarios that may materialize, from business as usual to the possibility of AGI. Aside from doing justice to the prevailing level of uncertainty, scenario planning makes the potential opportunities and risks tangible and helps us to develop contingency plans and be prepared for multiple possible outcomes.

Following are three technological scenarios spanning a wide range of possible outcomes that economic policymakers should pay attention to:

Scenario I (traditional, business as usual): Advances in AI boost productivity and automate a range of cognitive work tasks, but they also create new opportunities for affected workers to move into new jobs that are, on average, more productive than those from which they were displaced. This view is encapsulated by panel 1 of Chart 1.

Scenario II (baseline, AGI in 20 years): Over the next 20 years, AI gradually advances to the point of AGI, resulting in its ability to perform all human work tasks by the end of the period, devaluing labor (Susskind, forthcoming). This would correspond to the perspective of finite brainpower captured by panel 2 of Chart 1, together with the assumption that it would take 20 years for the most complex cognitive tasks to be accessible to AI.

Scenario III (aggressive, AGI in five years): This scenario replicates Scenario II but on a more aggressive timeline, such that AGI with all the associated consequences for labor would be reached within five years.

Although I am highly uncertain, at the time of writing, I estimate that each of these scenarios has a greater than 10 percent probability of materializing. To account for the uncertainty and adequately prepare for the future, I believe that policymakers should take each of these scenarios seriously, stress-test how our economic and financial policy frameworks would perform in each scenario, and where necessary reform them to ensure that they would be adequate.

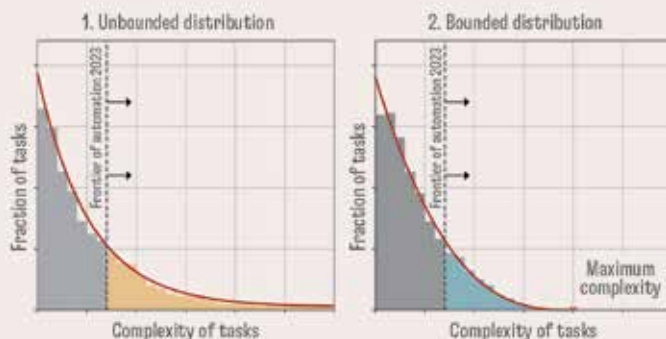
The three scenarios have the potential to lead to markedly different economic outcomes across a wide range of indicators, including economic growth, wages and returns to capital, fiscal sustainability, inequality, and political stability. Moreover, they call for reforms to our social safety nets and systems of taxation and affect the conduct of monetary policy, financial regulation, and industrial and development strategies.

Korinek and Suh (2023) analyze the implications of the scenarios described for output and wages in a

CHART 1

AI pushing the limits

Automation has already taken over many tasks that previously only humans could perform, as indicated by the gray bars. As AI advances, human beings may be able to continue to move into more complex tasks indefinitely (yellow bars), or AI may eventually overtake all human capabilities (blue bars).



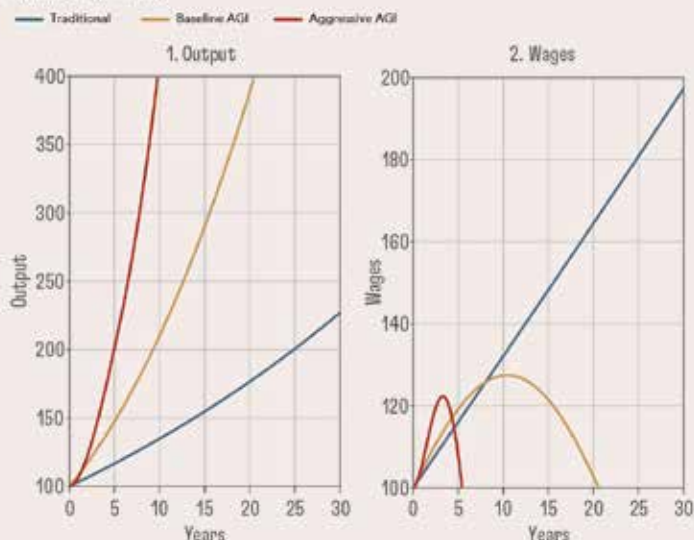
SOURCE: Anton Korinek.

CHART 2

Scenarios for output and wages

The advent of AGI leads to a takeoff in growth at the expense of plummeting competitive market wages.

(initial year = 100)



SOURCE: Anton Korinek.
NOTE: AGI = artificial general intelligence.

mainstream macroeconomic model of automation. The results for all three scenarios are illustrated in Chart 2, in which the path of output for each scenario is displayed in panel 1 and the path of competitive market wages in panel 2.

Three main insights stand out:

First, whereas growth continues along the trajectory we are used to from past decades in the conservative business-as-usual scenario, output growth in the two AGI scenarios is much faster, as the scarcity of labor is no longer a constraint on output.

Second, wages initially rise in all three scenarios—but only as long as labor is scarce. They plummet as the economy is close to reaching AGI.

Third, the takeoff in output and the collapse in wages in the two AGI scenarios are both driven by the same force: the substitution of scarce labor by comparatively more abundant machines. This suggests that it should be possible to design institutions that compensate workers for their income losses and ensure that the gains from AGI lead to shared prosperity.

Chart 2 illustrates the broad contours of how unprecedented technological changes may affect the macroeconomy, but it is best understood as an illustration of possibilities rather than as a precise prediction. A long list of caveats applies. First, the model underlying the chart is cast in an efficient economy in which labor earns competitive returns. A range of factors may slow the rollout of AGI compared with what is technologically possible, from organizational frictions, regulations, and constraints on capital accumulation—such as chip supply chain bottlenecks—to societal choices on the implementation of AGI. Even when it is technologically possible to replace workers, society may choose to keep humans in certain functions—for example, as priests, judges, or lawmakers. The resulting “nostalgic” jobs could sustain demand for human labor in perpetuity (Korinek and Juelfs, forthcoming).

To determine which AI scenario the future most resembles as events unfold, policymakers should monitor leading indicators across multiple domains, keeping in mind that all efforts to predict the pace of progress face tremendous uncertainty. Useful indicators span technological benchmarks, levels of investment flowing into AI development, adoption of AI technologies throughout the economy, and resulting macroeconomic and labor market trends. Technological benchmarks offer the most direct measure of how well AI systems perform a wide

“Given the starkly differing perspectives on future progress in AI, it would be unwise to put all eggs in one basket and formulate economic plans for a single scenario.”

range of labor tasks. Levels of investment, such as investment in research and development, talent, and computer chips, capture how much of our resources are flowing into AI development. Indicators of growing AI adoption through all sectors of the economy would capture whether the resulting systems are usefully deployed in practice. Finally, the macroeconomic implications would eventually become visible in productivity statistics and labor market trends.

Tracking these complementary signals allows policymakers to tailor policy responses to the realities of AI as they manifest. But we must remain humble—the future is likely to surprise us.

The starkly different economic trajectories implied by the three scenarios described earlier underscore the importance of developing adaptive policy frameworks that can respond nimbly as the future unfolds. Policymakers should stress-test existing institutions against each scenario and reform them where necessary to ensure they are resilient. This may involve gradual steps, such as reforming systems of taxation and expanding social safety nets, or new programs, such as introducing small basic incomes that can be scaled up when necessary.

Policymakers should charge teams of experts with iterative scenario planning to help them regularly update their views on how the probabilities of the various scenarios evolve. Embracing the uncertainty through an adaptable, scenario-based approach will allow us to maximize the benefits and mitigate the risks in the economic sphere from AI’s continuing evolution. **F&D**

ANTON KORINEK is a professor in the Department of Economics and Darden School of Business at the University of Virginia. He is economics of AI lead at the Centre for the Governance of AI, a research associate at the National Bureau of Economic Research, and a research fellow at the Centre for Economic Policy Research.

REFERENCES

- Korinek, Anton. 2023. Generative AI for Economic Research: Use Cases and Implications for Economists. *Journal of Economic Literature* 2023, 61(4).
- Korinek, Anton, and Megan Juelfs. Forthcoming. “Preparing for the (Non-existent?) Future of Work.” *Oxford Handbook of AI Governance*. Oxford, UK: Oxford University Press.
- Korinek, Anton, and Donghyun Suh. 2023. “Scenarios for the Transition to AGI.” University of Virginia working paper, Charlottesville, VA.
- Susskind, Daniel. Forthcoming. “Technological Unemployment.” *Oxford Handbook of AI Governance*. Oxford, UK: Oxford University Press.