The Economics of Transformative AI: A Research Agenda

Erik Brynjolfsson[1], Anton Korinek[2], and Ajay Agrawal[3]

Draft: Please do not quote or circulate. Comments welcome.

As AI capabilities rapidly advance, our understanding of the potential economic, social, and geopolitical impacts lags behind. To address this gap, we propose a research agenda for the economic implications of transformative artificial intelligence (TAI). Key areas of inquiry include economic growth, innovation, income distribution, decision-making power, geopolitics, information flows, AI safety, and human well-being. We discuss several methodological approaches, including theoretical modeling, new economic indicators, welfare metrics, task-level assessments, and AI-agent simulations. By accelerating research in these areas, economists can develop insights and tools to navigate the challenges and opportunities TAI presents. The paper emphasizes interdisciplinary collaboration and proactive policy development to ensure broad distribution of TAI's benefits while mitigating potential risks.

Acknowledgments: *We thank Basil Halperin and Philip Trammell for their excellent research assistance on this draft and the numerous participants in the March 2024 Asilomar Workshop on the Economics of Transformative AI for valuable discussions and suggestions. The Stanford Digital Economy Lab provided funding and support.*

[1] Stanford Digital Economy Lab and the Stanford Institute for Human-Centered AI, Stanford University.
[2] Department of Economics and Darden School of Business, University of Virginia, and GovAI.
[3] Rotman School of Management, University of Toronto.

1. Motivation

There has been extraordinary improvement in the capabilities of AI in recent years. Driven by significant investments and rapid advances in computing power, better algorithms, and more data, AI is poised to transform industries and reshape our economy and society. The rapid technological progress has exposed a critical disparity: our institutions, organizational structures, skills, and most importantly, our economic frameworks and models, are struggling to keep pace. In this growing gap lie the greatest risks and challenges of the coming decade, as well as the greatest opportunities. To address this gap, we must speed up our understanding of AI's economic implications.

Before we can formulate effective solutions, we must define the fundamental questions to be addressed in clear and simple terms. As the old adage suggests, "A problem well-stated is half-solved." Accordingly, this paper aims to identify and delineate these key questions, laying the groundwork for a productive research agenda. By doing so, we hope to catalyze the insights and analytical tools necessary to navigate the evolving landscape of AI. Ultimately, we can not only be better observers and analysts but also provide the insights needed to shape a future of broadly shared prosperity.

As rapid as improvements in AI have been recently, from acing exams in law, medicine, math, and science, to writing essays and creating art, there are strong reasons to believe that even bigger changes will occur within the next few years. For one thing, companies are rapidly increasing their investments. For instance, while OpenAI is expected to spend $5.4 billion on computing resources alone through 2024, that number is projected to increase over seven-fold by 2029.[4] Other companies are increasing their investments at a similar pace. The number of researchers working in the field is soaring. We are also beginning to see AI contribute to the research effort of improving itself, as it makes coding and the design of algorithms and chips more productive.

One marker of these advances has been the change in projected timelines for achieving various milestones in AI. For instance, Metaculus is a global forecasting platform that aggregates predictions from a diverse community of forecasters, using machine learning to optimize forecasts on key global issues. In 2020, the median prediction for a "general AI system" that could outperform most humans on a broad set of cognitive and physical tasks was the year 2062. Two years ago, the median prediction was 2042. As of this writing, the median is now 2032, while a quarter expect it to be achieved by 2027.[5] Dario Amodei, CEO of Anthropic, recently wrote that it might happen as soon as 2026.[6] Robotics is likely to lag behind cognitive

---

[4] Cade Metz, Mike Isaac, and Erin Griffith, "Microsoft and OpenAI's Close Partnership Shows Signs of Fraying," *The New York Times*, October 17, 2024, https://www.nytimes.com/2024/10/17/technology/microsoft-openai-partnership-deal.html

[5] Matthew Barnett, "When will the first general AI system be devised, tested, and publicly announced?," Metaculus, in progress, https://www.metaculus.com/questions/5121/date-of-artificial-general-intelligence/.

[6] Dario Amodei, "Machines of Loving Grace," October 2024, https://darioamodei.com/machines-of-loving-grace. Amodei uses a slightly different definition of "powerful AI" which may account for some of the difference in timelines.

tasks, with 2035 the median estimate on Metaculus for when "reliable and general household robot" will be developed.[7]

Impressive as these technological feats may be, the societal implications will be even more transformative. Even current AI technologies have the potential to affect large sectors of the economy. For instance, a recent analysis of the effect of GPT-4 on over 18,000 tasks estimates that 19% of jobs in the US economy will have at least half their tasks significantly affected, while 80% of jobs will have 10% or more of their tasks affected.[8] If, as expected, much more capable forms of AI are developed, the effect will be commensurately broader and deeper.

We believe that the need to understand the economic implications of transformative AI (TAI) is urgent. If machines that exhibit human or superhuman capabilities on a broad set of metrics are created, many of our existing economic institutions, norms, and systems will need to be fundamentally reinvented. As with other exponential processes, the societal changes are likely to unfold slowly at first but then suddenly. Thus, we cannot wait to address these questions once the technology is fully developed and the urgent need for fundamental economic transformation is upon us. By then, it could be too late to adequately mitigate the negative impacts, or fully realize the potential positive ones. Just as leading labs in companies, governments and universities are accelerating their investments into the technologies of AI, we need to accelerate research on the economic and societal implications.

2. What Transformative AI Means for the Economy

2.1 What Is Transformative AI?

There are ongoing debates about what constitutes transformative AI (TAI) and related concepts such as artificial general intelligence (AGI) and human-level AI (HLAI). TAI, as defined by Holden Karnofsky, refers to AI systems that precipitate a transition comparable to (or more significant than) the agricultural or industrial revolutions, but occurring over a much shorter time period.[9] This definition emphasizes the scale and speed of change rather than specific technical capabilities.

In contrast, the term AGI focuses more on the versatility and general problem-solving abilities of AI. It is often used to refer to AI systems that can perform any intellectual task that a human can. Similarly, HLAI is benchmarked against human cognitive abilities across a wide range of

---

[7] Matthew Barnett, "When will a reliable and general household robot be developed?," Metaculus, in progress, https://www.metaculus.com/questions/16625/date-of-reliable-and-general-household-robots/

[8] Tyna Eloundou et al., "GPTs are GPTs: Labor market impact potential of LLMs," *Science* 384, no. 6702 (2024): 1306-08, https://doi.org/10.1126/science.adj0998.

[9] Holden Karnofsky, "Some Background on Our Views Regarding Advanced Artificial Intelligence," Open Philanthropy, May 6, 2016, https://www.openphilanthropy.org/research/some-background-on-our-views-regarding-advanced-artificial-intelligence/.

tasks. It's worth noting that while TAI focuses more on outcomes, AGI and HLAI concentrate on the technology itself. We will systematize and generalize this approach in the following analysis.

Our focus for this paper is on the economic dimension of transformative AI. We aim to understand not only the economic impact of such systems but also the economic forces at work in their development and deployment. This approach allows us to highlight the role of our economic system and policies in shaping societal outcomes, much as they did during previous technological revolutions. Furthermore, the insights we gain may help us steer technological advances in a beneficial direction while mitigating foreseeable risks.

2.2 Economic Indicators for Transformative AI

Our perspective informs a set of technological and economic indicators relevant for analyzing the progress and impact of transformative AI.[10]

2.2.1. Indicators of factor inputs help us track the resources devoted to AI development and deployment, which reflect the scale and pace of advancement. They also provide insights into potential resource constraints or environmental impacts of AI progress.

a) **Use of compute and efficiency of compute use:** This set of indicators measures the total computational power used in frontier AI training and/or deployment, often quantified in floating-point operations per second (FLOPS) or petaflop/s-days. We should also track improvements in compute efficiency, such as the reduction in compute needed to achieve specific AI benchmarks over time.
b) **Labor inputs to AI and robotics:** This could measure the headcount or total compensation paid to personnel working on producing cutting-edge AI and robotics systems.
c) **Energy consumption:** This set of indicators measures the total energy used by AI systems, including both training and inference phases. It could be quantified in kilowatt-hours (kWh) or joules, and might be normalized per unit of compute or per AI task performed.
d) **Use of other raw materials:** This could include the consumption of other important input factors, such as materials crucial for AI and robotic hardware. Metrics might include the volume or value of these materials used in manufacturing.

2.2.2. Technological indicators, reflecting the economic definition of the term "production technology" as reflecting how we turn inputs into outputs, capture how AI is transforming the economy's production process, for example, by replacing—or possibly augmenting—human labor. They provide crucial insights into the potential for AI to transform the economy and the changing nature of work.

[10] There won't necessarily be a single point in any one of these indicators that would constitute a sharp boundary for TAI. Instead, a body of indicators can help us better comprehend the nature and rate of transformation.

a) **Advances in AI capabilities:** This has traditionally been measured through performance on standardized benchmarks across a wide range of domains (e.g., MMLU, BigBench, HumanEval, MATH). We must track both the absolute performance and the rate of improvement over time. Importantly, part of our role as economists is to examine to what extent these technical indicators map into economic usefulness.
b) **Advances in robot capabilities:** This includes metrics on dexterity, mobility, and task completion rates for physical robots. We might measure the percentage of human physical tasks that robots can perform or the speed and accuracy with which they complete standardized tasks. Again, for our purposes, the focus would be on economic usefulness.
c) **Substitutability of labor with capital:** This could be quantified through the elasticity of substitution measures between AI/robotic systems and human labor across different job categories, together with the elasticity of substitution of the resulting outputs in consumers' consumption baskets. This also includes the percentage of tasks within occupations that can be automated (i.e. perfectly substituted).

2.2.3. Production/output indicators gauge the economic impact of AI on overall production, efficiency, and the labor market. They are crucial for understanding how the benefits of AI are distributed across the economy and society.

a) **Productivity (including Total Factor Productivity (TFP) and labor productivity):** We should measure changes in output per unit of input, both for the economy as a whole (TFP) and specifically for labor. This could be tracked at the firm, industry, and economy-wide levels.
b) **Output growth:** This would measure the overall increase in economic output (e.g., GDP growth) attributable to AI adoption. We might develop methods to isolate AI's contribution to growth from other factors.
c) **Effects on labor demand:** have both a price and a quantity dimension:
    i) On the price side, labor demand is reflected in wages. We should track changes in wage levels across different skill categories and occupations, paying particular attention to those most affected by AI.
    ii) On the quantity side, labor demand is reflected in job numbers and flows, including job displacement and creation. Moreover, it may also be reflected in labor force participation rates or unemployment numbers. At the sectoral level, we can observe shifts in employment across sectors.

2.2.4. Financial market indicators often reflect expectations about future technological impacts, making these indicators valuable for anticipating economic shifts. They can provide early signals of how investors and businesses are valuing AI's potential.

a) **Equity markets:** We could track the stock performance of AI-focused companies and the adoption of AI in various sectors. This might include specialized AI stock indices or the AI-related revenue of major tech companies.

b) **Energy prices:** Given the energy-intensive nature of AI, tracking energy prices (especially electricity) could provide insights into the costs and constraints of AI deployment.
c) **Interest rates:** Changes in interest rates might reflect expectations about AI-driven productivity growth. We should analyze the relationship between AI advancement and long-term interest rate trends.

2.2.5. Industry-level phenomena reveal how AI is transforming the structure of the economy and creating new opportunities. They provide insights into the dynamism of the economy and the pace of creative destruction driven by AI.

a) **Emergence of new industries:** We should track the number and growth of new AI-enabled industries. This could include measures such as the number of new NAICS codes related to AI, or the revenue and employment in entirely new categories of business.
b) **Rapid industry reshuffling:** This could be measured through changes in market concentration (e.g., Herfindahl-Hirschman Index) within industries, the rate of company formations and bankruptcies, and shifts in industry compositions of major stock indices.

2.2.6. Income distribution and inequality indicators reflect the societal impacts as AI potentially reshapes the distribution of economic gains—these indicators become crucial for understanding societal impacts. They can help policymakers identify and address potential increases in inequality resulting from AI adoption.

a) **Labor share vs. capital share:** We should track the proportion of national income going to workers versus capital owners. This could be measured economy-wide and within specific sectors, especially those heavily impacted by AI.
b) **Gini coefficient:** This standard measure of income inequality should be monitored at various levels (within countries, between countries, and globally) to assess how AI affects overall income distribution.

2.2.7. International indicators help us understand how AI might affect global economic relationships and potentially exacerbate or reduce international inequalities. They are crucial for anticipating geopolitical implications and informing international economic policies.

a) **Global terms of trade:** We should monitor how AI affects the relative prices of exports and imports for different countries. This could include tracking changes in the value of knowledge-intensive exports relative to raw materials or manufacturing.
b) **Cross-country gaps in GDP/capita:** This would involve measuring how AI adoption affects economic convergence or divergence between countries at different levels of development.

It is important to note that our final objective is ultimately social welfare so the purpose of all these indicators is that they reflect different dimensions of AI's effects on welfare. Utilitarian welfare depends not only on individual utility derived from material consumption of goods and services but also on non-material goods such as health, happiness, and meaning. As we track these economic indicators, we must keep in mind the broader implications of AI-triggered transformation for human well-being.

2.3 Concrete Manifestations of Transformative AI

Experts have proposed a number of definitions and benchmarks for TAI. We can relate these to our different categories of economic indicators:

Manifestations in Technological Indicators

- One perspective focuses on specific milestones in the percentage of human tasks that can be automated by AI, for example, 80% of goods production or of ideas production. For example, Vinod Khosla proposes a working definition of AI being capable of performing 80% of human tasks across 80% of jobs,[11] A shortcoming of such manifestations is that the fraction of human tasks is itself endogenous. For instance, we have likely already automated more than 80% of the tasks that humans performed in 1800.
- A more demanding indicator would be for AI to fully automate all cognitive tasks—as in the "drop-in remote worker" scenario whereby an AI system could perform all the capabilities that could be performed by a remote worker.
- Dario Amodei defines "powerful AI" as AI systems that not only surpass human expertise in a wide range of fields but also understand and interact with the world in a sophisticated manner, enabling them to collaborate with humans and solve complex problems.[12]
- OpenAI defines five levels of AI capability in its roadmap towards AGI, starting from conversational AI and progressing through reasoners, agents, and innovators.[13] The final level envisions AI systems capable of functioning as entire organizations, representing a significant leap in autonomy and problem-solving abilities across various domains.

Manifestations in Production/Output Indicators

Another perspective focuses on how AI affects output indicators such as productivity, real gross domestic product (GDP), or real GDP/capita, where the term “real” implies adjusted for inflation.

- Some observers define threshold indicators for annual output growth or productivity growth—for example, one metric suggests that TAI would be marked by annual growth

[11] Khosla, Vinod. “AI: Dystopia or Utopia? Summary.” Khoslaventures.com, 2024. Accessed October 8, 2024. https://www.khoslaventures.com/ai-dystopia-or-utopia-summary/.
[12] Amodei, Dario. “Machines of Loving Grace.” Darioamodei.com. Accessed October 11, 2024. https://darioamodei.com/machines-of-loving-grace.
[13] “OpenAI’s 5 Steps to AGI.” Perplexity.ai, 2024. Accessed October 18, 2024. https://www.perplexity.ai/page/openai-s-5-steps-to-agi-STzklF5SSQ6JOiBTaV.cfA.

of real output of at least 30%, amounting to roughly a ten-fold increase in the growth rate from the average of the Industrial Age.[14]

- Mustafa Suleyman's concept of "Artificially Capable Intelligence" focuses on AI systems that could transform $10,000 into $1 million within a specific time frame.[15]

These concrete manifestations provide tangible benchmarks for identifying TAI, complementing the more general economic indicators discussed earlier.

One of the key objectives of our research agenda is to encourage further work on evaluating how well these and other indicators capture the economic variables most relevant to helping humanity manage the transition to TAI. As AI continues to advance, refining and expanding these indicators will be crucial for policymakers, economists, and society at large to anticipate, prepare for, and shape the economic impacts of this transformative technology. By advancing our understanding of these economic dimensions, we can better navigate the challenges and opportunities presented by TAI, ensuring that its development aligns with our societal goals and values.

3. Top economic questions and challenges

What are the big questions raised by TAI?

Undoubtedly there will be surprises and “unknown unknowns.” But it’s also clear that many of the issues that economists have grappled with will need to be fundamentally rethought. In particular, to prepare for the economic and societal impacts of transformative AI, we need to address the following questions, grouped into nine broad areas.

Economic Growth

1. *How will TAI change the determinants and rate of economic growth?*
2. *Which factors will be the main bottlenecks for output growth?*
3. *How will TAI affect the relative scarcity of factors, including labor, capital and compute?*
4. *How will the role of human capital change?*

Invention, Discovery and Innovation

5. *How will TAI boost the rate and direction of invention, discovery, and innovation?*

Income Distribution

---

[14] Open Philanthropy, “Report on Whether AI Could Drive Explosive Economic Growth | Open Philanthropy,” August 9, 2023, https://www.openphilanthropy.org/research/report-on-whether-ai-could-drive-explosive-economic-growth/.
[15] Mustafa Suleyman, “Mustafa Suleyman: My New Turing Test Would See if AI Can Make $1 Million,” *MIT Technology Review*, August 31, 2023, https://www.technologyreview.com/2023/07/14/1076296/mustafa-suleyman-my-new-turing-test-would-see-if-ai-can-make-1-million/.

6. *How will TAI affect labor markets, and how will this be reflected in wages and employment?*
7. *Will TAI exacerbate income and wealth inequality?*
8. *How will TAI interact with our social safety nets?*

Concentration of Decision-making and Power

9. *What are the risks of AI-driven economic power becoming concentrated in the hands of a few companies or countries?*
10. *How might AI shift political power dynamics?*

Geopolitics

11. *How could AI redefine the structure of international relations, including trade, global security, economic power and inequality, political stability, and global governance?*

Information, Communication, and Knowledge

12. *How will truth vs. misinformation, cooperation vs. polarization, and creativity and insight vs. confusion be amplified or dampened?*
13. *How will AI affect the spread of information and knowledge, particularly regarding the distinction between real and synthetic data?*

AI Safety & Alignment

14. *How can we balance the economic benefits of TAI with its risks, including potentially existential risks?*
15. *What can economists contribute to help align TAI with social preferences and welfare?*

Meaning and Well-being

16. *How can people retain their sense of meaning and worth if "the economic problem is solved" as Keynes predicted?*

While these areas of inquiry are already complex and challenging, there is another crucial dimension to consider across all of them: understanding and successfully managing the transition from our current economic institutions, organizations, and processes to those shaped by TAI. This transition phase is particularly critical given the potential for rapid, non-linear and even non-monotonic changes in many important metrics in the years ahead. The success of the transition will likely determine how well societies can harness the benefits of TAI while minimizing its risks and disruptions.

3.1 Economic Growth

The arrival of TAI would represent a paradigm shift in the drivers of economic growth. As AI systems approach the ability to perfectly substitute for labor, standard growth models predict that the rate of economic growth will greatly rise.[16] This raises questions about how TAI will alter

[16] Philippe Aghion, Benjamin F. Jones, and Charles I. Jones, "Artificial Intelligence and Economic Growth," in *National Bureau of Economic Research*, ed. Ajay Agrawal, Joshua Gans, and Avi Goldfarb, *The*

the traditional determinants of economic growth, such as productivity, capital accumulation, and technological progress. How can economists develop methods to detect early signs of an AI-driven "growth explosion," improving upon existing approaches like those of Nordhaus?[17]

As TAI reshapes the economy, the nature of growth constraints is likely to change significantly. Current bottlenecks, such as limited supply of workers, human capital, and physical capital, may evolve or be supplanted by new limiting factors. Which factors will emerge as the main bottlenecks for economic growth in a TAI-driven economy, and how will these differ from traditional constraints? Will energy availability, computational resources, or raw materials become key constraints? How might these changing bottlenecks affect the distribution of economic gains across different sectors and populations?

The role of human capital in the economy is likely to undergo significant changes as AI capabilities progress. Some forms of human capital may depreciate rapidly. How will the value and role of human capital evolve as TAI is developed? Which human skills are likely to remain in demand as TAI capabilities grow, and how might this evolve over time? How should education and training systems adapt to prepare workers for an economy where TAI is prevalent? What are the implications of the changing valuation of human capital for labor markets?

The pattern and speed of TAI adoption across different sectors and countries will significantly influence its economic impact. This diffusion process will be crucial in determining how the potential benefits of TAI are realized and distributed. What factors will influence the rate of TAI adoption and diffusion across sectors and countries, and how will these affect economic growth? Might differences in TAI diffusion contribute to economic divergence between sectors or countries? What policies could promote optimal diffusion of TAI to maximize economic benefits while minimizing potential negative consequences?

3.2 Invention, Discovery and Innovation

Innovation is the primary driver of economic growth. It is the main source of productivity gains and enables the development of new markets. The continuous accumulation of technological progress thus forms the basis for long-run economic growth, as outlined in endogenous growth theories, where innovation drives labor and capital productivity gains which in turn enable more innovation.

How will TAI transform the innovation process? To what extent will TAI automate scientific discovery? Hypothesis generation? Hypothesis testing? Traditional innovation processes are costly and also are constrained by the time required to formulate hypotheses, conduct experiments, and iterate on solutions. TAI systems, with their capacity to analyze complex data

---

Economics of Artificial Intelligence: An Agenda (University of Chicago Press, 2019), 237-282, https://www.nber.org/system/files/chapters/c14015/c14015.pdf.

[17] William D. Nordhaus, “Are We Approaching an Economic Singularity? Information Technology and the Future of Economic Growth,” *American Economic Journal: Macroeconomics* 13, no. 1 (2021): 299-332, https://doi.org/10.1257/mac.20170105.

and autonomously generate insights, may dramatically reduce the costs and time involved in these activities. How will the ability to automate experimentation and problem-solving at scale influence the rate of technological progress? What are the likely bottlenecks? What will be the impact on the frequency and quality of innovations, and, consequently, the rate of economic growth?

In addition to making the innovation process faster and cheaper, TAI may also transform the process from disciplinary to non-disciplinary. Many of the most transformative innovations emerge from the integration of knowledge across distinct fields, yet human cognitive limitations often prevent experts from identifying novel intersections between disciplines. A vast fraction of the discovery-based institutions (e.g., universities and institutes) today are largely organized around disciplines, including training, promotion, and dissemination. To what extent will TAI, equipped with access to vast cross-disciplinary data and capable of synthesizing disparate knowledge, identify previously unrecognized complementarities? By generating insights that bridge multiple domains, how might TAI impact not only the rate of innovation but also the direction by shifting the innovation process from discovering local maxima (disciplinary discoveries) to more global discoveries (not constrained by disciplines)? What are the implications for economic growth?

To what extent will TAI transform the innovation process by democratizing access? Today, innovation is primarily concentrated in firms and institutions with substantial resources and expertise, which limits the participation of smaller actors. Will TAI provide small firms, entrepreneurs, and individuals with enhanced capacity to perform sophisticated innovation activities, including R&D, previously out of reach? Although humans will still need to articulate their desired outcome, to what extent will TAI be able to do most of the rest of the work of the innovation process? To what extent will TAI shift the distribution of the agents that engage in the innovation process? How will this impact the rate and direction of innovation? What are the implications for productivity and economic growth?

3.3 Income Distribution

Labor is the main source of income for the majority of the population in modern societies, and labor markets therefore play a crucial role in income distribution. Technological progress has traditionally gone hand in hand with both job displacement and the creation of new work, although the two have not always been in balance.[18] However, advances in AI may be far more rapid and transformative than earlier technologies.[19] As we observed earlier, concepts like AGI or HLAI explicitly focus on AI systems being able to perform most or all human cognitive tasks. Simple economics suggests that if a machine can perform a worker's job, her wage will fall to the machine's cost.

---

[18] David Autor et al., "New Frontiers: The Origins and Content of New York, 1940-2018," August 14, 2022, https://economics.mit.edu/sites/default/files/2022-11/ACSS-NewFrontiers-20220814.pdf.

[19] Anton Korinek and Joseph E. Stiglitz, "Artificial Intelligence and Its Implications for Income Distribution and Unemployment," in *National Bureau of Economic Research*, ed. Ajay Agrawal, Joshua Gans, and Avi Goldfarb, *The Economics of Artificial Intelligence: An Agenda* (University of Chicago Press, 2019), 349-390, https://www.nber.org/system/files/chapters/c14018/c14018.pdf.

How will TAI affect labor markets? Will the capabilities of TAI largely displace workers across the economy, or will there be pockets or sectors with significant labor demand in a world of TAI? What types of jobs will remain if machines can perform essentially all cognitive tasks? These may include both jobs that remain human for transitional reasons and jobs in which human labor is demanded for intrinsic human reasons.[20] What will the balance between cognitive and physical work look like under TAI? What will be the ultimate implications for equilibrium wages and employment levels as well as unemployment rates?

What will be the balance between labor and capital, and how will the labor share of income evolve? Investments in compute, energy, and related factors may grow in relative importance compared to labor. Moreover, the purveyors of TAI are likely to earn significant windfall gains from their inventions. Will this lead to an increase in income inequality in relative terms? Or will the growing economy lift all boats and ensure that the remaining labor demand is sufficient to spread around the surplus generated by TAI somewhat equitably? And how will these developments affect the concentration and inequality of wealth?

Our existing social safety nets are also designed around the idea that labor is the most important source of income. Social security and health benefits depend on people's jobs or labor income. There is disability insurance for people who can't work, and unemployment insurance for people who lose their jobs. Will our social safety nets still perform their role effectively in a TAI world? How can they be reformed to be optimally adapted to a world of TAI? What other mechanisms of social insurance and income distribution can we design to ensure that TAI's benefits are shared broadly across society?

3.4 Concentration of Decision-making and Power

A core question in economics has long been the allocation of decision rights and power, from the socialist calculation debate,[21] to analyzing the boundaries of the firm,[22] market structure, and corporate decision-making. Friedrich Hayek argued that the detailed knowledge needed for economic decision-making was inherently dispersed.[23] For instance, a local shopkeeper understands the specific needs and preferences of their customers, a skilled carpenter possesses knowledge about the properties of their tools, and a farmer possesses unique knowledge about the soil conditions. Much of this knowledge has historically been tacit and not subject to aggregation. What's more, no single human brain can fully understand all the steps needed to create even a product as simple as a pencil, from refining each of its raw materials to manufacturing and marketing it economically. These information-processing constraints have

[20] Anton Korinek, "Economic Policy Challenges for the Age of AI," *National Bureau of Economic Research*, September 2024, https://doi.org/10.3386/w32980.
[21] David M. Levy and Sandra J. Peart, "Socialist Calculation Debate," in *The New Palgrave Dictionary of Economics*, December 13, 2016, https://doi.org/10.1057/978-1-349-95121-5_2070-1
[22] Oliver Hart, "An Economist's Perspective on the Theory of the Firm," *Columbia Law Review* 89, no. 7 (1989): 1757-1774.
[23] Friedrich Hayek, "The Use of Knowledge in Society," *The American Economic Review* 35, no. 4 (September 1945): 519–530, https://www.kysq.org/docs/Hayek_45.pdf.

been important determinants of the degree of centralization of decision-making, the boundaries between markets and firms, and the role of government in directing economic activity.

As AI, and digital technologies more broadly, change the cost of information processing by orders of magnitude, and as machine learning enables the rapid discovery of previously tacit knowledge, it would be surprising if economic organizations and institutions did not also change. For instance, will larger retailers gain an increased competitive advantage over small single-unit shops? Will similar dynamics increase concentration in other industries? Or will AI democratize expertise and lead to flourishing competition?

Relatedly, the success of ever-larger models in line with AI scaling laws suggests the possibility that the AI industry itself may become increasingly concentrated, while the rapid cost declines for models of nearly-equivalent performance, and the success of open source models could support increased competition.

The questions raised are not merely whether economic decision-making and power will become more centralized or decentralized, but what types of decisions will be most affected and the implications for markets, firms, governments, and new forms of economic organization.

Furthermore, economic concentration often leads to concentration of political power. So changes in the locus of economic decision-making may ripple through to other parts of the economy. In particular, even if a highly centralized system could optimize more types of economic decision-making, it might at the same time erode individual autonomy and liberty. Thus, understanding the effects on economic and political power is one of the key questions raised by more powerful AI.

3.5 Geopolitics

TAI may have profound implications for geopolitics by influencing military capabilities, transforming international competition and trade, and altering global governance frameworks. TAI's potential capacity to accelerate technological development, disrupt labor markets, and enhance decision-making may create new competitive advantages for states, potentially redefining the structure of international relations. Key geopolitical impacts include (1) the military applications of TAI and their effect on global security, (2) shifts in economic power due to productivity gains that are not uniformly distributed across states, (3) international governance and regulatory frameworks, (4) potential inequality between technologically advanced and lagging states, and (5) challenges to political stability and information control within countries. Thus, there is a rich research agenda in this domain.

Military Applications: The incorporation of TAI into military operations may alter global security dynamics by enabling faster decision-making, autonomous weapons, and enhanced intelligence and surveillance capabilities. How will TAI reshape the economics of deterrence and the balance of power among states? How will TAI impact the economics of military alliance stability and strategic rivalries? To what extent will autonomous weapons systems lower the threshold for conflict, or, conversely, create new avenues for deterrence? How will TAI change the economics of cyber warfare and the defense of critical infrastructure? What are the economics of TAI in military contexts in terms of legal and ethical challenges, such as implications for international humanitarian law and the regulation of autonomous weapons systems? Finally, what regulatory frameworks are needed to manage the dual-use nature of AI technologies without hindering economic growth?

Uneven Productivity Gains and Trade: Under what conditions will TAI-driven productivity gains exacerbate global economic inequalities by favoring early adopters and technologically advanced countries? To what extent will factors such as existing industrial capacity, research infrastructure, or policy frameworks determine which countries capture the most value from TAI? How might TAI reshape global trade patterns, particularly with respect to the competitive position of emerging economies? Under what conditions will states adopt certain strategies to mitigate labor displacement and avoid social instability caused by rapid automation? Under what conditions will TAI alter the geopolitical importance of industries such as manufacturing, agriculture, and services, and how might trade policies evolve in response to these transformations?

Technologically Lagging States: What policies can countries with limited technological capacity adopt to avoid marginalization in a TAI-driven global economy? How might global inequality affect the economic and political influence of technologically lagging states in international institutions? What role can foreign aid or technology-sharing initiatives play in closing the TAI gap between countries? How will TAI-induced inequality affect migration patterns, regional stability, and global cooperation? How will TAI influence cooperation on shared challenges like climate change or public health crises?

Global Governance and Regulatory Frameworks: The development of international governance frameworks for TAI will be critical to managing geopolitical risks and ensuring the technology evolves peacefully and benefits all countries. Under what conditions will international institutions balance the sovereignty of nation-states with the need for coordinated TAI governance? What models of governance—whether centralized or decentralized—might prove most effective for regulating TAI? How can the global community address concerns around regulatory arbitrage, where countries adopt divergent rules to gain competitive advantages? How will regulatory capture concerns influence the role of major technology companies in shaping regulatory standards?

Political Stability and Information Control: How will TAI affect political stability by enhancing the ability of both states and non-state actors to manipulate information? What are the implications of TAI for the economics of authoritarian regimes, democratic governance, and social cohesion? How might TAI tools be used to influence the economics of information with respect to shaping public opinion, disrupting political processes, or suppressing dissent? What economic policies can protect electoral integrity and combat disinformation? How will the widespread use of TAI influence the economics of state surveillance practices and the balance between national security and civil liberties? How might TAI-driven disruptions to labor markets contribute to political instability and social unrest, particularly in regions with weak institutional frameworks? Understanding these dynamics will help policymakers seeking to harness the benefits of TAI while mitigating its risks to political stability.

3.6 Information, Communication, and Knowledge

One of the key determinants of the economic success of a society is how it manages information, communication, and knowledge. Laws, institutions, incentives, and norms that promote the creation and transmission of accurate information will tend to boost economic growth. For instance, the U.S. Constitution gives Congress the power to promote the progress of science and useful arts by granting authors and inventors exclusive rights to their writings and discoveries for a limited time. More broadly, the jury system, blind scientific reviews, libel laws, and the scientific method itself are systems that, often imperfectly, are intended to differentially favor truthful information.

Digital information systems, communications, and social media are increasingly intermediated by AI moderators, filters, and amplifiers and are increasingly populated by AI agents. In some contexts, these systems have been found to disproportionately promote misinformation, perhaps unintentionally.[24]

Relatedly, AI systems are creating content themselves. This content can be useful and innovative. It may provide deeper and broader insights and even novelty. It can also be misleading and destructive. AI-based deepfakes can be designed to misrepresent people and events. AI can simply overwhelm the audience with the sheer quantities of content.

No doubt, the design choices of these socio-technical systems will affect the outcomes, raising a fruitful set of research questions and opportunities.

3.7 AI Safety & Alignment

AI safety and alignment refer to the challenge of ensuring that artificial intelligence systems behave in ways that are consistent with human values and intentions.[25] As AI systems become more powerful and autonomous, the economic implications of their safety and alignment become increasingly important. How can we define and measure AI safety and alignment both in economic terms and in the economic realm? What are the economic incentives for developing safe and aligned AI systems? How do the costs of ensuring AI safety and alignment compare to the potential economic benefits of AI development?

---

[24] Soroush Vosoughi, Deb Roy, and Sinan Aral, "The spread of true and false news online," *Science* 359, no. 6380 (2018): 1146-1151, https://doi.org/10.1126/science.aap9559.
[25] AI Safety Summit, commissioned by the United Kingdom, "International Scientific Report on the Safety of Advanced AI," https://www.gov.uk/government/publications/international-scientific-report-on-the-safety-of-advanced-ai.

Economists can play a crucial role in developing frameworks and methodologies to align AI systems with social preferences and welfare.[26] We have a powerful toolkit to describe the relationships between a wide set of variables and social welfare, and we also have experience in designing mechanisms for aligning agents with their principals. The questions we face include: How can social welfare functions be adapted or extended to capture the complexities of AI alignment? - What economic mechanisms can be designed to internalize the externalities (both positive and negative) generated by AI systems? How can we design incentive structures that encourage AI developers and users to prioritize alignment with broader societal goals? What role can market mechanisms play in promoting the development of safe and aligned AI systems? What economic tools can be developed to address the challenges of preference aggregation and value learning in AI systems?

One particular trade-off has been analyzed by Chad Jones:[27] TAI systems may significantly increase economic growth but also pose existential risks to humanity. How can we quantify the potential economic benefits of TAI in terms of productivity growth, innovation, and overall economic output? What methods can economists develop to assess and model the existential risks associated with TAI systems? What are appropriate economic frameworks for risk-benefit analysis that incorporate both the potential for near-infinite economic growth and the possibility of human extinction? Under what conditions is it rational to continue rapid AI progress, and under what conditions should development be slowed or halted? And Jones (2024) also suggests factoring the potential improvements in human longevity and mortality reduction from AI into benefit-risk calculations?

An important factor to consider is the risk of AI race dynamics, whereby actors who benefit from being the first to develop higher capabilities prioritize speed over safety since the safety risks of AI are borne by a wider set of actors. For example, within labs, individual researchers may perceive career benefits to advancing more rapidly. Within nations, labs are racing against each other to be the first to ship more powerful capabilities. And at the geopolitical scale, as emphasized above, individual countries race to outdo each other. In a worst-case scenario, race dynamics may give rise to the materialization of existential risks. This is a classic externality problem, and economists have significant experience both in analyzing and internalizing externalities.

3.8 Meaning and Well-being

As we approach a potential future where TAI may reduce the role of human labor, questions of meaning and well-being become increasingly crucial. Keynes' prediction about solving the "economic problem" raises fundamental questions about human purpose and fulfillment in a TAI world. How can economics contribute to our understanding of meaning and well-being in a world

[26] Anton Korinek and Avital Balwit, "Aligned with Whom? Direct and Social Goals for AI Systems," *The Oxford Handbook of AI Governance* (2023): 65-85, https://doi.org/10.1093/oxfordhb/9780197579329.013.4.

[27] Charles I. Jones, "The A.I. Dilemma: Growth versus Existential Risk," *National Bureau of Economic Research* (2024), https://web.stanford.edu/~chadj/existentialrisk.pdf.

without work? What frameworks can we develop to analyze the production and distribution of non-monetary sources of fulfillment? Importantly, what is our final objective in a world where machines can perform essentially all work, and would it even be desirable for work to maintain its current societal importance if we achieve TAI?

The psychological and social impacts of widespread unemployment present a complex picture. Interestingly, studies show that retirees often experience increased happiness and life satisfaction, while the involuntarily unemployed tend to suffer decreased well-being. This apparent contradiction raises several questions: What factors contribute to the positive experience of retirees versus the negative experience of the unemployed? How do societal expectations, financial security, and the voluntary nature of retirement influence these outcomes? Can we design economic policies that mimic the positive aspects of retirement for those displaced by TAI?

In a world where the role of traditional employment as a primary source of purpose and identity declines, individuals and society will need to redefine these concepts. How can economics help model the transition from a work-centric society to one where meaning is derived from other activities? What role might education, creative pursuits, community engagement, or leisure play in providing structure and identity? How can we quantify and analyze the value of these non-monetary benefits of work in economic terms?

Many of the non-monetary aspects of work may be subject to externalities or internalities that must be considered in the policy debate—otherwise, it can just be left to individuals to decide whether they want to work at the prevailing market wage.[28] Externalities arise when an individual's work affects others in society beyond just producing marketable output, such as by fostering social connections or political stability. Internalities occur when individuals don't fully internalize the effects of their work choices on their own welfare. How can we accurately measure and model these externalities and internalities in economic frameworks? What are the implications for optimal labor allocation and social welfare if these factors are significant? How might the nature and magnitude of work-related externalities and internalities change in a world where TAI significantly reduces the need for human labor? What policy interventions might be justified to address these market failures, and how would they differ from current labor market policies?

As we potentially transition to a post-work society, economics will play a crucial role in understanding and shaping new institutions to support human flourishing. What economic frameworks can help us analyze the "meaning production" in a society where TAI handles most economic tasks? How might the distribution of meaning-generating activities be optimized for social welfare? What role could TAI itself play in creating or facilitating new sources of meaning and fulfillment? How can we ensure that the benefits of a TAI economy are equitably distributed, not just in terms of material wealth, but also in terms of access to fulfilling activities and meaning?

---

[28] Anton Korinek and Megan Juelfs, "Preparing for the (Non-Existent?) Future of Work," *The Oxford Handbook of AI Governance* (2024): 746-776, https://doi.org/10.1093/oxfordhb/9780197579329.013.44.

4. Methodologies for the Economics of TAI

Economics offers a rich toolkit to analyze social science questions. We briefly lay out the main methodologies that we envision to make progress on the research agenda we have outlined above.

4.1 Theoretical Approaches

Theoretical approaches play a crucial role in understanding the potential economic impacts of TAI, especially given the unique challenges posed by this emerging technology. Since TAI may represent a radical break from the past, one of the primary difficulties researchers face is the scarcity of relevant historical data. This lack of precedent means that a significant amount of work in this field involves predicting a future that is still largely unclear. In this context, theoretical approaches that take a step back and leverage higher-level regularities from the past—such as fundamental laws of economics—become particularly valuable in attempting to glimpse the future economic landscape shaped by TAI.

These theoretical approaches encompass a range of modeling techniques designed to understand potential economic shifts. Growth models, including those focusing on AI take-off dynamics, are at the forefront of this research. These models aim to capture the potentially explosive and non-linear growth patterns that TAI might induce, helping to forecast scenarios of rapid technological advancement and its economic consequences. Complementing these macro-level models are micro-to-macro approaches that examine the economic effects of AI. These approaches bridge the gap between individual-level impacts of AI adoption and their aggregate effects on the broader economy, providing a more comprehensive understanding of how TAI might reshape economic structures and interactions.

A special focus within theoretical approaches is on normative frameworks, particularly those drawing from public finance principles. These frameworks address critical questions about the societal implications of TAI, such as how to equitably share the benefits of AI and how to reform taxation systems in a world where traditional labor may be significantly diminished. A point that is particularly salient is how to steer AI development in socially beneficial directions.[29][30] By tackling these normative questions, theoretical approaches not only help predict the future economic landscape but also provide valuable insights for policymakers and society at large in navigating the transformative effects of AI.

4.2 A Transformative AI Dashboard

Extensive work is already been done to track current AI capabilities directly (see e.g. HAI's AI Index Report 2024, Ch. 2). Economic indicators of a coming AI-driven boom in economic growth

---

[29] Anton Korinek and Joseph E. Stiglitz, "Steering Technological Progress," *National Bureau of Economic Research* (2020), https://conference.nber.org/conf_papers/f143989.pdf.
[30] Erik Brynjolfsson, "The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence," January 11, 2022, https://doi.org/10.48550/arXiv.2201.04200.

would also be valuable because the relationships between benchmark task scores and economic impacts are not always intuitive. Nordhaus (2021) takes a useful first step in this direction, looking for signs that computing is substituting for labor well enough to drive a near-term growth explosion. As of the dates of his article, he largely does not find them. That's not to say that the evidence will change as the technology improves.[31]

Work of this kind could fruitfully be extended to many other economic indicators, such as the substitution of labor for capital across various links in the AI and robotics supply chains. Further insights could also be developed in each domain by studying trends in a broader range of measures of substitutability: e.g. the elasticity of substitution between labor and capital instead of the capital share. These data could then be fit to a variety of growth models, reflecting differing assumptions about hard-to-measure variables, such as the relationship between research breakthroughs and economic growth on a given time horizon. Finally, by keeping the associated data and model-based forecasts current with an easily accessible "economic transformation dashboard," other researchers and policymakers could easily stay up to date about whether a period of AI-driven explosive growth appears to be approaching.

4.3 New Metrics for Welfare

New metrics are also needed for assessing economic welfare and consumption. For instance, as AI and other digital technologies become responsible for a larger share of production and distribution, the marginal costs of many goods and services will fall to nearly zero. The standard national accounts are based on GDP, but when goods have zero price, they often have zero weight in GDP. Likewise, labor productivity is typically measured as GDP per hour worked, so mismeasurement of GDP will ripple through to productivity measures as well.

New methods, such as massive online choice experiments, can help assess the valuations that consumers have for goods and services that are poorly captured by traditional measures.[32] As these methods are extended, refined, and scaled, they can create an updated measurement toolkit to better track AI's contributions in the coming years.

4.4 Task-level Assessments of Potential Impact

It has also proven fruitful to analyze the effects of AI and related technologies at the task level rather than the level of entire occupations, firms, or industries. This method has been applied to

---

[31] It's worth noting that Nordhaus's model focuses heavily on aggregate macroeconomic variables (such as the capital share). More "microfounded" work of this kind includes that of Besiroglu et al. (2023), who find a rising capital share in AI R&D in particular, suggesting that, in effect, machines may soon improve machine capabilities without being bottlenecked by a lack of human research capacity. On the other hand, Acemoglu (2024) argues that automating only the particular tasks AI is most clearly on track to automate would have little impact on growth.

[32] Erik Brynjolfsson and Avinash Collins, "How Should We Measure the Digital Economy?," Hutchins Center on Fiscal & Monetary Policy at Brookings (2020), https://www.brookings.edu/wp-content/uploads/2020/01/WP57-Collis_Brynjolfsson_updated.pdf

understand the potential effects of machine learning[33,34] and generative AI[35] on work and employment. While prior work used the Bureau of Labor Statistics O-Net taxonomy of about 18,000 tasks, future work could apply natural language processing to classify hundreds of millions of job postings and resumés, creating a much more fine-grained and dynamic task taxonomy.

4.5 Simulating Economies Using AI Agents

A fascinating new approach in economic research leverages the power of artificial intelligence itself, particularly through simulations and agent-based modeling. By creating agents based on large language models (LLM Agents), researchers can simulate human behavior with increasing accuracy and scale. Imagine the potential of modeling thousands, even millions, of these agents interacting in a simulated economy. The simulations could rapidly run the equivalent of randomized controlled trials (RCTs) to assess alternative policy options and interventions. This opens exciting avenues for improving our understanding of the economy and exploring how AI might reshape labor markets, consumer behavior, industrial growth, and the emergence of unforeseen economic bottlenecks.

This research frontier extends beyond mere simulation. As AI agents become more sophisticated, they will increasingly interact with each other and with humans in the real world. These interactions, potentially far more complex than those among humans alone, could reshape the foundations of economic activity. Understanding an economy where AI agents negotiate contracts, make investment decisions, or even drive consumer trends. This dynamic interplay between AI and human agents is poised to redefine the economic landscape in ways we are only beginning to grasp.

4.6 Simulations for Robotics

We are likely to achieve some form of TAI in the digital world long before we do so in the physical world. This is due to the paucity of physical world data for training AIs. The most commonly cited solution is simulation. Simulation would reduce the cost of generating TAI in the physical world by orders of magnitude. The primary application of TAIs in the physical world is control systems for robots. At present, robots trained using simulation data struggle to perform effectively in the real world due to the reality gap—a mismatch between simulated environments and real-world conditions. Simulations simplify physical interactions and often fail to capture complex forces, environmental variability, and material properties, leading to inaccuracies in robot performance. Additionally, simulated sensor data tends to be cleaner and more predictable

---

[33] Erik Brynjolfsson, Tom Mitchell, and Daniel Rock, "What Can Machines Learn, and What Does It Mean for Occupations and the Economy?," *AEA Papers and Proceedings* 108 (2018): 43-47, https://doi.org/10.1257/pandp.20181019.

[34] Erik Brynjolfsson and Tom Mitchell, "What can machine learning do? Workforce implications," *Science* 358, no. 6370 (2017): 1530-1534, https://doi.org/10.1126/science.aap8062

[35] Eloundou et al., "GPTs are GPTs: Labor market impact potential of LLMs," 1306-08.

than real-world inputs, which are affected by noise and environmental factors, impairing robots' perception capabilities. Simulations also expose robots to a limited range of scenarios, leaving them unprepared for unexpected real-world events.

What are the economic trade-offs, incentives, and policy implications associated with the development and deployment of simulation-based AI systems? What are the externalities associated with simulation systems and thus the welfare implications of creating these as a public good (e.g., open source)? How do the costs and benefits of using simulations to train AI for robots compare to traditional data collection methods, particularly in industries with high physical-world uncertainty? What are the optimal strategies for investing in simulation technologies to reduce the reality gap and improve robot performance in the real world? How will firms that adopt simulation-based AI systems early gain competitive advantages, and what risks will late adopters face if simulation technologies remain imperfect? How will investments in sim-to-real transfer learning or domain randomization impact productivity and labor markets, particularly in sectors reliant on automation? Finally, from a social welfare perspective, what is the role of public policy in encouraging innovation, such as incentives for collaborative simulation platforms or regulations to address externalities from poorly performing AI systems that cross the simulation-reality divide?

5. Conclusion

The transition to an economy shaped by TAI will not follow a predetermined path—there are multiple potential equilibria that could emerge. Some scenarios offer the promise of enhanced prosperity, where TAI drives unprecedented productivity, enhances social welfare, and distributes benefits reasonably fairly. However, without thoughtful management, other outcomes could be dystopian, with increased inequality, mass unemployment, and social instability, leaving many worse off than before.

The goal of this research agenda is to equip economic policymakers with the insights and tools necessary to shape policies that maximize the likelihood of positive outcomes. By identifying key economic indicators, anticipating challenges, and advancing this research agenda, we aim to increase the likelihood that the introduction of TAI will lead to shared prosperity and a sustainable future for humanity.

As the research agenda outlined in this paper progresses, it will naturally give rise to important policy considerations that must be addressed. Key areas may include: reforming labor laws to account for AI-driven changes in employment; adapting taxation systems to ensure equitable distribution of AI-generated wealth; revamping education and skill development programs to prepare the workforce for an AI-driven economy; strengthening social insurance and income distribution mechanisms; maintaining social and political stability in the face of rapid technological change; adjusting macroeconomic policy frameworks to account for AI's impact on productivity and growth; updating antitrust and market regulations to address AI-driven market concentration; refining intellectual property frameworks to balance innovation incentives with societal benefits; developing environmental and resource policies that account for AI's ecological impact; and establishing global AI governance structures to ensure responsible development and deployment of TAI. By addressing these policy areas proactively, we can work towards harnessing the full potential of TAI while mitigating its risks and ensuring its benefits are broadly shared across society.

There is a critical need for interdisciplinary research that integrates insights from economics, technology, public policy, and other social sciences. While significant resources are being allocated to the *technical* development of TAI, there has been comparatively little investment in understanding its *economic* implications or preparing policies for its arrival. Immediate action is essential to address these gaps, with the next steps involving articulating a more comprehensive research agenda[36], establishing working groups, striking partnerships across sectors, and securing institutional support to guide research efforts. A proactive approach will be crucial to ensure that TAI enhances human welfare rather than exacerbating existing inequalities.

Preparing for the arrival of TAI demands more than technical breakthroughs—it requires robust policy frameworks to manage disruptions and shape outcomes aligned with societal goals. Only by accelerating research on the economics of TAI and coordinating global efforts can we create the tools and policies needed to navigate this transformation. With timely attention to this topic, we can seize the opportunities that TAI offers while mitigating the risks, increasing the likelihood of a future of broadly shared prosperity and human flourishing.

[36] This is only a first draft!