The Turing Trap: The Promise & Peril of Human-Like Artificial Intelligence

Erik Brynjolfsson

In 1950, Alan Turing proposed a test of whether a machine was intelligent: could a machine imitate a human so well that its answers to questions were indistinguishable from a human's? Ever since, creating intelligence that matches human intelligence has implicitly or explicitly been the goal of thousands of researchers, engineers, and entrepreneurs. The benefits of human-like artificial intelligence (HLAI) include soaring productivity, increased leisure, and perhaps most profoundly a better understanding of our own minds. But not all types of AI are human-like – in fact, many of the most powerful systems are very different from humans – and an excessive focus on developing and deploying HLAI can lead us into a trap. As machines become better substitutes for human labor, workers lose economic and political bargaining power and become increasingly dependent on those who control the technology. In contrast, when AI is focused on augmenting humans rather than mimicking them, humans retain the power to insist on a share of the value created. What is more, augmentation creates new capabilities and new products and services, ultimately generating far more value than merely human-like AI. While both types of AI can be enormously beneficial, there are currently excess incentives for automation rather than augmentation among technologists, business executives, and policy-makers.

Alan Turing was far from the first to imagine human-like machines.[1] According to legend, 3,500 years ago, Dædalus constructed humanoid statues that were so lifelike that they moved and spoke by themselves.[2] Nearly every culture has its own stories of human-like machines, from Yanshi's leather man described in the ancient Chinese *Liezi* text to the bronze Talus of the Argonautica and the towering clay *Mokkerkalfe* of Norse mythology. The word robot first appeared in Karel Čapek's influential play *Rossum's Universal Robots* and derives from the Czech word *robota*, meaning servitude or work. In fact, in the first drafts of his play, Čapek named them *labori* until his brother Josef suggested substituting the word robot.[3]

Of course, it is one thing to tell tales about humanoid machines. It is something else to create robots that do real work. For all our ancestors' inspiring stories, we are the first generation to build and deploy real robots in large numbers.[4]

Dozens of companies are working on robots as human-like, if not more so, as those described in the ancient texts. One might say that technology has advanced sufficiently to become indistinguishable from mythology.[5]

The breakthroughs in robotics depend not merely on more dexterous mechanical hands and legs, and more perceptive synthetic eyes and ears, but also on increasingly human-like artificial intelligence ( ). Powerful systems are crossing key thresholds: matching humans in a growing number of fundamental tasks such as image recognition and speech recognition, with applications from autonomous vehicles and medical diagnosis to inventory management and product recommendations.[6]

These breakthroughs are both fascinating and exhilarating. They also have profound economic implications. Just as earlier general-purpose technologies like the steam engine and electricity catalyzed a restructuring of the economy, our own economy is increasingly transformed by . A good case can be made that is the most general of all general-purpose technologies: after all, if we can solve the puzzle of intelligence, it would help solve many of the other problems in the world. And we are making remarkable progress. In the coming decade, machine intelligence will become increasingly powerful and pervasive. We can expect record wealth creation as a result.

Replicating human capabilities is valuable not only because of its practical potential for reducing the need for human labor, but also because it can help us build more robust and flexible forms of intelligence. Whereas domain-specific technologies can often make rapid progress on narrow tasks, they founder when unexpected problems or unusual circumstances arise. That is where human-like intelligence excels. In addition, could help us understand more about ourselves. We appreciate and comprehend the human mind better when we work to create an artificial one.

These are all important opportunities, but in this essay, I will focus on the ways that could lead to a realignment of economic and political power.

The distributive effects of depend on whether it is primarily used to augment human labor or automate it. When augments human capabilities, enabling people to do things they never could before, then humans and machines are complements. Complementarity implies that people remain indispensable for value creation and retain bargaining power in labor markets and in political decision-making. In contrast, when replicates and automates existing human capabilities, machines become better substitutes for human labor and workers lose economic and political bargaining power. Entrepreneurs and executives who have access to machines with capabilities that replicate those of humans for a given task can and often will replace humans in those tasks.

Automation increases productivity. Moreover, there are many tasks that are dangerous, dull, or dirty, and those are often the first to be automated. As more

tasks are automated, a fully automated economy could, in principle, be structured to redistribute the benefits from production widely, even to those people who are no longer strictly necessary for value creation. However, the beneficiaries would be in a weak bargaining position to prevent a change in the distribution that left them with little or nothing. Their incomes would depend on the decisions of those in control of the technology. This opens the door to increased concentration of wealth and power.

This highlights the promise and the peril of achieving : building machines designed to pass the Turing Test and other, more sophisticated metrics of human-like intelligence.[7] On the one hand, it is a path to unprecedented wealth, increased leisure, robust intelligence, and even a better understanding of ourselves. On the other hand, if leads machines to automate rather than augment hu-

able to solve types of problems that are solvable by any existing human, animal, or machine. That suggests that is not human-like.

The good news is that both automation and augmentation can boost labor productivity: that is, the ratio of value-added output to labor-hours worked. As productivity increases, so do average incomes and living standards, as do our capabilities for addressing challenges from climate change and poverty to health care and longevity. Mathematically, if the human labor used for a given output declines toward zero, then labor productivity would grow to infinity.[10]

The bad news is that no economic law ensures everyone will share this growing

If capital in the form of can perform more tasks, those with unique assets, talents, or skills that are *not* easily replaced with technology stand to benefit disproportionately.[18] The result has been greater wealth concentration.[19]

Ultimately, a focus on more human-like can make technology a better substitute for the many nonsuperstar workers, driving down their market wages, even as it amplifies the market power of a few.

To understand the limits of substitution-oriented automation, consider a thought experiment. Imagine that our old friend Dædalus had at his disposal an extremely talented team of engineers 3,500 years ago and built human-like machines that

required: fully 60 percent of people are now employed in occupations that did not exist in 1940.[38] In short, automating labor ultimately unlocks less value than augmenting it to create something new.

At the same time, automating a whole job is often brutally difficult. Every job involves multiple different tasks, including some that are extremely challenging to automate, even with the cleverest technologies. For example, may be able to read mammograms better than a human radiologist, but it is not very good at the other twenty-six tasks associated with the job, according to - , such as comforting a concerned patient or coordinating on a care plan with other doctors.[39]

income of 37 percent, while long capital gains have a variety of favorable rules, including a lower statutory tax rate of 20 percent, the deferral of taxes until capital gains are realized, and the "step-up basis" rule that resets capital gains to zero, wiping out the associated taxes, when assets are inherited.

The first rule of tax policy is simple: you tend to get less of whatever you tax. Thus, a tax code that treats income that uses labor less favorably than income derived from capital will favor automation over augmentation. Treating both business models equally would lead to more balanced incentives. In fact, given the positive externalities of more widely shared prosperity, a case could be made for treating wage income *more* favorably than capital income, for instance by expanding the earned income tax credit.[44] It is unlikely that any government official can define in advance exactly which technologies and innovations augment humans rather than merely substitute for them; indeed, most technologies have elements of each and the outcome depends a great deal on how they are deployed. Thus, rather than prescribe or proscribe specific technologies, a broad-based set of incentives can gently nudge technologists and managers toward augmentation on the margin, much as carbon taxes encourage myriad types of cleaner energy or research and development tax credits encourage greater investments in research.

Government policy in other areas could also do more to steer the economy clear

-

A

A A

Erik Brynjolfsson *Machine, Platform, Crowd: Harnessing Our Digital Future* (2017), *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies* (2014), *Race against the Machine: How the Digital Revolution Is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy* (2011) *Wired for Innovation: How Information Technology Is Reshaping the Economy* (200)

1 *Mind* 5 (236) 433 460, /10.10 3/ .236.433. *The Discourse,*

2 1 , 201 , 1 .04 .000 2 1&

3 22, 2011,

4

2 2 0 (, 2020).

5

6 2021. (2021), . 2, //. /2103.06312.

, – *International Journal of Computer Vision* 115 (3) (2015) 211 252.

, – *Harvard Business Review* (2017) 3 11.

7 , *What Computers Can't Do* (, 1 72) ! – *AI Magazine* 26 (4) (2005) 6 , – *AI Magazine* 37 (1) (2016) 3 4.

– 6 .

, – *Behavioral and Brain Sciences* 3 (3) (1 0) 417 457.

10 ( )

11 , – *The Quarterly Journal of Economics* 70 (1) (1 56) 65 4.

12 , – *Review of Economic Studies* 6 (4) (2002) 7 1 0 .

13 , *Race Against the Machine: How the Digital Revolution Is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy* (, 2011) , – *American Economic Review* 10 (6) (201) 14 1542.

14 1 20 167 2000, , ., ., *How Was Life? Global Well-Being since 1820* (, 2014).

15 , 30, 2021. (–).

16 , – *The Quarterly Journal of Economics* 12 (1) (2014) 61 103 , , , – . 255 (, , 201). y, , ,

2 … 24, 2015, … // … (… 2021).

30 … *Journal of Political Economy* … 4 (4) … 6 … 71 … *Journal of Political Economy* … (6) … 111 … 115.

31 … 2021).

32 … y 4, 2021, … //…

33 … *Mind Children: The Future of Robot and Human Intelligence* (… 1…).

34 … *Technology Review,* … 2017.

35 … *Financial Times,* … y 12, 201 .

36 … *MIT Sloan Management Review* 62 (1) (2020) 1 5.

37 … *The Economics of New Goods* (… 1… 6).

3 … 1… 40 201 … y 26, 2021.

3 … *Nature Reviews Clinical Oncology* 17 (134) (2020), … //… /10.103 /… 41571-020-032…-7.

40 … *AEA Papers and Proceedings* (201) 43 47.

41 … *Governance in an Emerging New World* (61…) (201…), … //…

42 … *American Economic Review* 0 (3) (1… 0) 511 52 .

43 … *Brookings Papers on Economic Activity* (… 2020) … *Redesigning AI* (… 2021).

44 … *Econometrica* 54 (3) (1… 6) 607 622 … *Journal of Public Economics* 2 (1) (1… 5) 5… 3.

45 [illegible]

46 [illegible], *Chutes and Ladders: Navigating the Low-Wage Labor Market* ([illegible], 2006).

47 [illegible]