

# WILL A.I. BECOME THE NEW MCKINSEY?

*As it's currently imagined, the technology promises to concentrate wealth and disempower workers. Is an alternative possible?*

**By Ted Chiang**

May 4, 2023

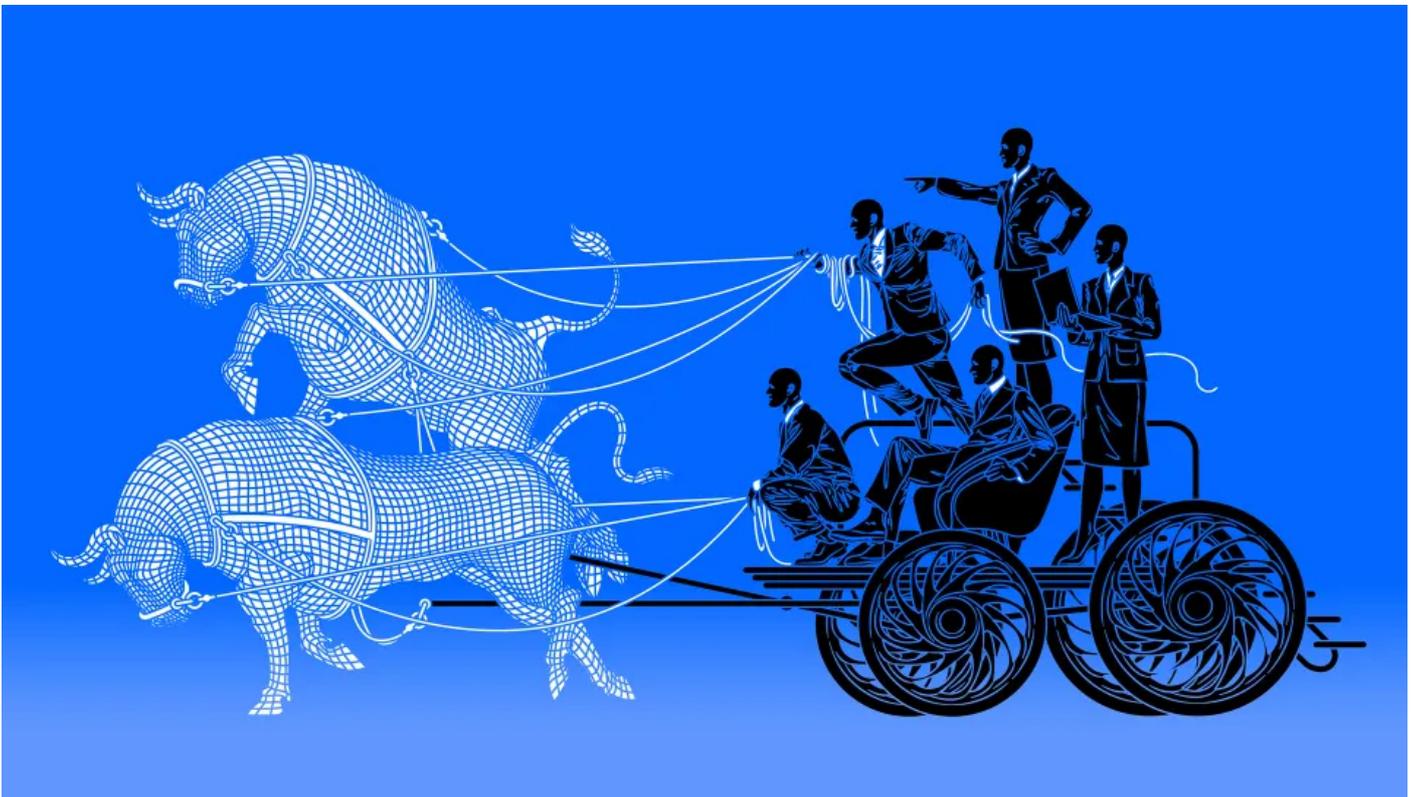


Illustration by Berke Yazicioglu



Save this story

**W**hen we talk about artificial intelligence, we rely on metaphor, as we always do when dealing with something new and unfamiliar. Metaphors are, by their nature, imperfect, but we still need to choose them

carefully, because bad ones can lead us astray. For example, it's become very common to compare powerful A.I.s to genies in fairy tales. The metaphor is meant to highlight the difficulty of making powerful entities obey your commands; the computer scientist Stuart Russell has cited the parable of King Midas, who demanded that everything he touched turn into gold, to illustrate the dangers of an A.I. doing what you tell it to do instead of what you want it to do. There are multiple problems with this metaphor, but one of them is that it derives the wrong lessons from the tale to which it refers. The point of the Midas parable is that greed will destroy you, and that the pursuit of wealth will cost you everything that is truly important. If your reading of the parable is that, when you are granted a wish by the gods, you should phrase your wish very, very carefully, then you have missed the point.

So, I would like to propose another metaphor for the risks of artificial intelligence. I suggest that we think about A.I. as a management-consulting firm, along the lines of McKinsey & Company. Firms like McKinsey are hired for a wide variety of reasons, and A.I. systems are used for many reasons, too. But the similarities between McKinsey—a consulting firm that works with ninety per cent of the Fortune 100—and A.I. are also clear. Social-media companies use machine learning to keep users glued to their feeds. In a similar way, Purdue Pharma used McKinsey to figure out how to “turbocharge” sales of OxyContin during the opioid epidemic. Just as A.I. promises to offer managers a cheap replacement for human workers, so McKinsey and similar firms helped normalize the practice of mass layoffs as a way of increasing stock prices and executive compensation, contributing to the destruction of the middle class in America.

A former McKinsey employee has described the company as “capital's willing executioners”: if you want something done but don't want to get your hands dirty, McKinsey will do it for you. That escape from accountability is one of the most valuable services that management consultancies provide. Bosses have certain goals, but don't want to be blamed for doing what's necessary to

achieve those goals; by hiring consultants, management can say that they were just following independent, expert advice. Even in its current rudimentary form, A.I. has become a way for a company to evade responsibility by saying that it's just doing what "the algorithm" says, even though it was the company that commissioned the algorithm in the first place.

The question we should be asking is: as A.I. becomes more powerful and flexible, is there any way to keep it from being another version of McKinsey? The question is worth considering across different meanings of the term "A.I." If you think of A.I. as a broad set of technologies being marketed to companies to help them cut their costs, the question becomes: how do we keep those technologies from working as "capital's willing executioners"? Alternatively, if you imagine A.I. as a semi-autonomous software program that solves problems that humans ask it to solve, the question is then: how do we prevent that software from assisting corporations in ways that make people's lives worse? Suppose you've built a semi-autonomous A.I. that's entirely obedient to humans—one that repeatedly checks to make sure it hasn't misinterpreted the instructions it has received. This is the dream of many A.I. researchers. Yet such software could easily still cause as much harm as McKinsey has.

Note that you cannot simply say that you will build A.I. that only offers pro-social solutions to the problems you ask it to solve. That's the equivalent of saying that you can defuse the threat of McKinsey by starting a consulting firm that only offers such solutions. The reality is that Fortune 100 companies will hire McKinsey instead of your pro-social firm, because McKinsey's solutions will increase shareholder value more than your firm's solutions will. It will always be possible to build A.I. that pursues shareholder value above all else, and most companies will prefer to use that A.I. instead of one constrained by your principles.

Is there a way for A.I. to do something other than sharpen the knife blade of

capitalism? Just to be clear, when I refer to capitalism, I'm not talking about the exchange of goods or services for prices determined by a market, which is a property of many economic systems. When I refer to capitalism, I'm talking about a specific relationship between capital and labor, in which private individuals who have money are able to profit off the effort of others. So, in the context of this discussion, whenever I criticize capitalism, I'm not criticizing the idea of selling things; I'm criticizing the idea that people who have lots of money get to wield power over people who actually work. And, more specifically, I'm criticizing the ever-growing concentration of wealth among an ever-smaller number of people, which may or may not be an intrinsic property of capitalism but which absolutely characterizes capitalism as it is practiced today.

As it is currently deployed, A.I. often amounts to an effort to analyze a task that human beings perform and figure out a way to replace the human being. Coincidentally, this is exactly the type of problem that management wants solved. As a result, A.I. assists capital at the expense of labor. There isn't really anything like a labor-consulting firm that furthers the interests of workers. Is it possible for A.I. to take on that role? Can A.I. do anything to assist workers instead of management?

Some might say that it's not the job of A.I. to oppose capitalism. That may be true, but it's not the job of A.I. to strengthen capitalism, either. Yet that is what it currently does. If we cannot come up with ways for A.I. to reduce the concentration of wealth, then I'd say it's hard to argue that A.I. is a neutral technology, let alone a beneficial one.

**M**any people think that A.I. will create more unemployment, and bring up universal basic income, or U.B.I., as a solution to that problem. In general, I like the idea of universal basic income; however, over time, I've become skeptical about the way that people who work in A.I. suggest U.B.I. as a response to A.I.-driven unemployment. It would be different if we already had universal basic income, but we don't, so expressing support for it

seems like a way for the people developing A.I. to pass the buck to the government. In effect, they are intensifying the problems that capitalism creates with the expectation that, when those problems become bad enough, the government will have no choice but to step in. As a strategy for making the world a better place, this seems dubious.

You may remember that, in the run-up to the 2016 election, the actress Susan Sarandon—who was a fervent supporter of Bernie Sanders—said that voting for Donald Trump would be better than voting for Hillary Clinton because it would bring about the revolution more quickly. I don't know how deeply Sarandon had thought this through, but the Slovenian philosopher Slavoj Žižek said the same thing, and I'm pretty sure he had given a lot of thought to the matter. He argued that Trump's election would be such a shock to the system that it would bring about change.

What Žižek advocated for is an example of an idea in political philosophy known as accelerationism. There are a lot of different versions of accelerationism, but the common thread uniting left-wing accelerationists is the notion that the only way to make things better is to make things worse. Accelerationism says that it's futile to try to oppose or reform capitalism; instead, we have to exacerbate capitalism's worst tendencies until the entire system breaks down. The only way to move beyond capitalism is to stomp on the gas pedal of neoliberalism until the engine explodes.

I suppose this is one way to bring about a better world, but, if it's the approach that the A.I. industry is adopting, I want to make sure everyone is clear about what they're working toward. By building A.I. to do jobs previously performed by people, A.I. researchers are increasing the concentration of wealth to such extreme levels that the only way to avoid societal collapse is for the government to step in. Intentionally or not, this is very similar to voting for Trump with the goal of bringing about a better world. And the rise of Trump illustrates the risks of pursuing accelerationism as a strategy: things can get very bad, and stay very bad for a long time,

before they get better. In fact, you have no idea of how long it will take for things to get better; all you can be sure of is that there will be significant pain and suffering in the short and medium term.

I'm not very convinced by claims that A.I. poses a danger to humanity because it might develop goals of its own and prevent us from turning it off. However, I do think that A.I. is dangerous inasmuch as it increases the power of capitalism. The doomsday scenario is not a manufacturing A.I. transforming the entire planet into paper clips, as one famous thought experiment has imagined. It's A.I.-supercharged corporations destroying the environment and the working class in their pursuit of shareholder value. Capitalism is the machine that will do whatever it takes to prevent us from turning it off, and the most successful weapon in its arsenal has been its campaign to prevent us from considering any alternatives.

People who criticize new technologies are sometimes called Luddites, but it's helpful to clarify what the Luddites actually wanted. The main thing they were protesting was the fact that their wages were falling at the same time that factory owners' profits were increasing, along with food prices. They were also protesting unsafe working conditions, the use of child labor, and the sale of shoddy goods that discredited the entire textile industry. The Luddites did not indiscriminately destroy machines; if a machine's owner paid his workers well, they left it alone. The Luddites were not anti-technology; what they wanted was economic justice. They destroyed machinery as a way to get factory owners' attention. The fact that the word "Luddite" is now used as an insult, a way of calling someone irrational and ignorant, is a result of a smear campaign by the forces of capital.

Whenever anyone accuses anyone else of being a Luddite, it's worth asking, is the person being accused actually against technology? Or are they in favor of economic justice? And is the person making the accusation actually in favor of improving people's lives? Or are they just trying to increase the private accumulation of capital?

Today, we find ourselves in a situation in which technology has become conflated with capitalism, which has in turn become conflated with the very notion of progress. If you try to criticize capitalism, you are accused of opposing both technology and progress. But what does progress even mean, if it doesn't include better lives for people who work? What is the point of greater efficiency, if the money being saved isn't going anywhere except into shareholders' bank accounts? We should all strive to be Luddites, because we should all be more concerned with economic justice than with increasing the private accumulation of capital. We need to be able to criticize harmful uses of technology—and those include uses that benefit shareholders over workers—without being described as opponents of technology.

Imagine an idealized future, a hundred years from now, in which no one is forced to work at any job they dislike, and everyone can spend their time on whatever they find most personally fulfilling. Obviously it's hard to see how we'd get there from here. But now consider two possible scenarios for the next few decades. In one, management and the forces of capital are even more powerful than they are now. In the other, labor is more powerful than it is now. Which one of these seems more likely to get us closer to that idealized future? And, as it's currently deployed, which one is A.I. pushing us toward?

Of course, there is the argument that new technology improves our standard of living in the long term, which makes up for the unemployment that it creates in the short term. This argument carried weight for much of the post-Industrial Revolution period, but it has lost its force in the past half century. In the United States, per-capita G.D.P. has almost doubled since 1980, while the median household income has lagged far behind. That period covers the information-technology revolution. This means that the economic value created by the personal computer and the Internet has mostly served to increase the wealth of the top one per cent of the top one per cent, instead of raising the standard of living for U.S. citizens

as a whole.

Of course, we all have the Internet now, and the Internet is amazing. But real-estate prices, college tuition, and health-care costs have all risen faster than inflation. In 1980, it was common to support a family on a single income; now it's rare. So, how much progress have we really made in the past forty years? Sure, shopping online is fast and easy, and streaming movies at home is cool, but I think a lot of people would willingly trade those conveniences for the ability to own their own homes, send their kids to college without running up lifelong debt, and go to the hospital without falling into bankruptcy. It's not technology's fault that the median income hasn't kept pace with per-capita G.D.P.; it's mostly the fault of Ronald Reagan and Milton Friedman. But some responsibility also falls on the management policies of C.E.O.s like Jack Welch, who ran General Electric between 1981 and 2001, as well as on consulting firms like McKinsey. I'm not blaming the personal computer for the rise in wealth inequality—I'm just saying that the claim that better technology will necessarily improve people's standard of living is no longer credible.

The fact that personal computers didn't raise the median income is particularly relevant when thinking about the possible benefits of A.I. It's often suggested that researchers should focus on ways that A.I. can increase individual workers' productivity rather than replace them; this is referred to as the augmentation path, as opposed to the automation path. That's a worthy goal, but, by itself, it won't improve people's economic fortunes. The productivity software that ran on personal computers was a perfect example of augmentation rather than automation: word-processing programs replaced typewriters rather than typists, and spreadsheet programs replaced paper spreadsheets rather than accountants. But the increased personal productivity brought about by the personal computer wasn't matched by an increased standard of living.

The only way that technology can boost the standard of living is if there are

economic policies in place to distribute the benefits of technology appropriately. We haven't had those policies for the past forty years, and, unless we get them, there is no reason to think that forthcoming advances in A.I. will raise the median income, even if we're able to devise ways for it to augment individual workers. A.I. will certainly reduce labor costs and increase profits for corporations, but that is entirely different from improving our standard of living.

It would be convenient if we could assume that a utopian future is right around the corner and develop technology for use in that future. But the fact that a given technology would be helpful in a utopia does not imply that it's helpful now. In a utopia where there's a machine that converts toxic waste into food, generating toxic waste wouldn't be a problem, but, in the here and now, no one could claim that generating toxic waste is harmless.

Accelerationists might argue that generating more toxic waste will motivate the invention of a waste-to-food converter, but how convincing is that? We evaluate the environmental impact of technologies in the context of the mitigations that are currently available, not in the context of hypothetical future mitigations. By the same token, we can't evaluate A.I. by imagining how helpful it will be in a world with U.B.I.; we have to evaluate it in light of the existing imbalance between capital and labor, and, in that context, A.I. is a threat because of the way it assists capital.

**A** former partner at McKinsey defended the company's actions by saying, "We don't do policy. We do execution." But this is a pretty thin excuse; harmful policy decisions are more likely to be made when consulting firms—or new technologies—offer ways to implement them. The version of A.I. that's currently being developed makes it easier for companies to lay people off. So is there any way to develop a kind of A.I. that makes it harder?

In his book "How to Be an Anticapitalist in the 21st Century," the sociologist Erik Olin Wright offers a taxonomy of strategies for responding

to the harms of capitalism. Two of the strategies he mentions are smashing capitalism and dismantling capitalism, which probably fall outside the scope of this discussion. The ones that are more relevant here are taming capitalism and resisting capitalism. Roughly speaking, taming capitalism means government regulation, and resisting capitalism means grassroots activism and labor unions. Are there ways for A.I. to strengthen those things? Is there a way for A.I. to empower labor unions or worker-owned coöperatives?

In 1976, the workers at the Lucas Aerospace Corporation in Birmingham, England, were facing layoffs because of cuts in defense spending. In response, the shop stewards produced a document known as the Lucas Plan, which described a hundred and fifty “socially useful products,” ranging from dialysis machines to wind turbines and hybrid engines for cars, that the workforce could build with its existing skills and equipment rather than being laid off. The management at Lucas Aerospace rejected the proposal, but it remains a notable modern example of workers trying to steer capitalism in a more human direction. Surely something similar must be possible with modern computing technology.

Does capitalism have to be as harmful as it currently is? Maybe not. The three decades following the Second World War are sometimes known as the golden age of capitalism. This period was partially the result of better government policies, but the government didn't create the golden age on its own: corporate culture was different during this era. In General Electric's annual report from 1953, the company bragged about how much it paid in taxes and how much it was spending on payroll. It explicitly said that “maximizing employment security is a prime company goal.” The founder of Johnson & Johnson said that the company's responsibility to its employees was higher than its responsibility to its shareholders. Corporations then had a radically different conception of their role in society compared with corporations today.

Is there a way to get back to those values? It seems unlikely, but remember

that the golden age of capitalism came after the enormous wealth inequality of the Gilded Age. Right now we're living in a second Gilded Age, in which wealth inequality is about the same as it was back in 1913, so it's not impossible that we could go from where we are now to a second golden age. Of course, in between the first Gilded Age and the golden age we had the Great Depression and two World Wars. An accelerationist might say that those events were necessary to bring about the golden age, but I think most of us would prefer to skip over those steps. The task before us is to imagine ways for technology to move us toward a golden age without bringing about another Great Depression first.

We all live in a capitalist system, so we are all participants in capitalism whether we like it or not. And it's reasonable to wonder if there's anything you as an individual can do. If you work as a food scientist at Frito-Lay and your job is to invent new flavors of potato chip, I'm not going to say that you have an ethical obligation to quit because you're assisting the engine of consumerism. You're using your training as a food scientist to provide customers with a pleasant experience; that's a perfectly reasonable way to make a living.

But many of the people who work in A.I. regard it as more important than inventing new flavors of potato chip. They say it's a world-changing technology. If that's the case, then they have a duty to find ways for A.I. to make the world better without first making it worse. Can A.I. ameliorate the inequities of our world other than by pushing us to the brink of societal collapse? If A.I. is as powerful a tool as its proponents claim, they should be able to find other uses for it besides intensifying the ruthlessness of capital.

If there is any lesson that we should take from stories about genies granting wishes, it's that the desire to get something without effort is the real problem. Think about the story of "The Sorcerer's Apprentice," in which the apprentice casts a spell to make broomsticks carry water but is unable to make them stop. The lesson of that story is not that magic is impossible to

control: at the end of the story, the sorcerer comes back and immediately fixes the mess the apprentice made. The lesson is that you can't get out of doing the hard work. The apprentice wanted to avoid his chores, and looking for a shortcut was what got him into trouble.

The tendency to think of A.I. as a magical problem solver is indicative of a desire to avoid the hard work that building a better world requires. That hard work will involve things like addressing wealth inequality and taming capitalism. For technologists, the hardest work of all—the task that they most want to avoid—will be questioning the assumption that more technology is always better, and the belief that they can continue with business as usual and everything will simply work itself out. No one enjoys thinking about their complicity in the injustices of the world, but it is imperative that the people who are building world-shaking technologies engage in this kind of critical self-examination. It's their willingness to look unflinchingly at their own role in the system that will determine whether A.I. leads to a better world or a worse one. ♦