ASSOCIATION FOR GLOBAL DEBATE

2017 FALL SEASON

RESEARCH PACKET

NHSDLC WeChat

TABLE OF CONTENTS

GENERAL NOTES

A Note on this Guide:

You are not limited to the contents of this packet. Other sources of information, and other arguments, are available and useful. Do not limit your thinking to the arguments found here, and try not to limit your research to these sources. We welcome any comments, questions, or suggestions concerning this packet (or anything else we do) sent to contact@agdebate.com. Good luck!

For each article, we have included the title, author, publication of origin, and a link to the article. The articles have been edited for the sake of length and clarity; you are free to read the whole articles if you wish. Places where we have removed content are marked with square brackets and ellipses "[..]".

We have added our own notes before most articles. These generally provide a summary of the source, and will indicate the key things to look for when reading the article.

Sources used here reflect a variety of different viewpoints, and perspectives. This is normal. Keep in mind that in a developing rapidly like artificial intelligence, different companies and individuals might have disagreements on the potential harms or benefits. Articles will also be from different dates; government policies can vary from country to country, and may have changed since the publication of a certain article. It is not necessary to read this research packet all the way through. An important skill to develop when doing research in an academic setting is deciding what information and what topics are important, and reading about those more in depth.

The articles have been grouped into categories that cover different issues. However, you will often find many different applications of AI mentioned in one article. So, you may, for example, be able to find information about transportation and AI, in an article about government policy. If you are looking for a specific term or word, you can, and should, browse this document using control+F to search for that specific term.

Evidence and Debate

An argument is made up of a claim, a warrant, and an impact. It might not be difficult to think of an argument, but it can be more difficult to provide some reasons why that argument is true. A warrant can come from an academic source, such as a website or article. If you are researching warrants for your case, you can use an excerpt from the article that is making the argument you are trying to make, rather than summarizing that part of the article in your own words.

Keep in mind that there will be situations where the other team wants to see your evidence. It would be helpful to have on hand an easy way to show them your evidence. While many teams might choose to hand over their laptop, consider that you can also transfer the files via flash drive, or have a couple of pages printed out that you can give to the other team. This can be

helpful in saving time when they want to see your evidence, because you can quickly provide the other team with a one or two paragraph excerpt that they can read.

Once you have the excerpts and evidence that you want for your constructive, you will need to provide citations for your evidence. Citations are important because they provide information about your source, and let everyone know where you are getting your evidence from. All evidence needs citations because any evidence that is presented in the round needs to verifiable by your opponents or the judge. A citation should include at the minimum the article's title, the date the article was published, the author, and the author's relevant qualifications. You should check the qualifications of the author for your evidence to make sure that they are a credible source.

You do not have to use a piece of evidence in every argument you make, if what you say is a piece of common knowledge, or your point is backed up by sound logic. In these cases, evidence is not always necessary. For example, if you claim that "As AI increases efficiency, companies won't need as many employees and this will lead to unemployment" this is something that anyone who reads a newspaper would have heard about, so you don't necessarily need evidence, but it might be helpful to explain why this is the case. In addition, a citation from an expert on the size of the unemployment problem, or a statistic about the speed of the coming change can help emphasize why your claim is correct

It is important to make sure that you use good evidence. The evidence you use should be; recent, from a qualified or relevant source, and sound in its reasoning and methodology. It is important to cite evidence that is recent. How recent your evidence should be can vary based on the subject. For example, an article about an ongoing diplomatic crisis might be out of date even if it is only a few weeks old. On the other hand, an article about government investment, can be several years old and still very relevant. When choosing your sources, please avoid websites like Wikipedia or Baidu Baike as these online encyclopedias can be edited by anyone, so their accuracy cannot be guaranteed. Examples of good sources are; newspapers, magazines, academic publications, and information from organizations that do research into relevant issues such as think tanks. Keep in mind that sources might have a certain political bias or agenda that they are trying to promote.

DEFINITIONS

What exactly constitutes AI can be a little difficult to pin down as you look through sources, since the precise definition of artificial intelligence (AI) is disputed among researchers in the technology industry. Even though the understanding of what exactly constitutes an AI can vary, what is important for this debate topic is to understand that generally, artificial intelligence systems are guided by the principles of flexibility, the ability to learn new tasks (often called machine learning), and the use of neural networks. Below are a couple of excerpts from the technology industry, as well as an encyclopedia definition, to help give you an understanding of AI.

Artificial Intelligence:

B.J. Copeland, *Encyclopaedia Britannica,* January 12,2017
https://www.britannica.com/technology/artificial-intelligence

Editors' Note:

This definition provides a short and concise understanding of AI, as well as a basic outline of the history of artificial intelligence and where we are in the development of a general AI.

~

Artificial intelligence, the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings. The term is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience. Since the development of the digital computer in the 1940s, it has been demonstrated that computers can be programmed to carry out very complex tasks—as, for example, discovering proofs for mathematical theorems or playing chess—with great proficiency. Still, despite continuing advances in computer processing speed and memory capacity, there are as yet no programs that can match human flexibility over wider domains or in tasks requiring much everyday knowledge. On the other hand, some programs have attained the performance levels of human experts and professionals in performing certain specific tasks, so that artificial intelligence in this limited sense is found in applications as diverse as medical diagnosis, computer search engines, and voice or handwriting recognition. […]

The Hype—and Hope—of Artificial Intelligence

Om Malik, *The New Yorker* Aug 26, 2016.
http://www.newyorker.com/business/currency/the-hype-and-hope-of-artificial-intelligence

Editors' Note:
In this excerpt, a former I.B.M employee gives us her understanding of what constitutes intelligence by dividing intelligence into three "stages" of development.
~

Michelle Zhou spent over a decade and a half at I.B.M. Research and I.B.M. Watson Group before leaving to become a co-founder of Juji, a sentiment-analysis start-up. An expert in a field where artificial intelligence and human-computer interaction intersect, Zhou breaks down A.I. into three stages. The first is recognition intelligence, in which algorithms running on ever more powerful computers can recognize patterns and glean topics from blocks of text, or perhaps even derive the meaning of a whole document from a few sentences. The second stage is cognitive intelligence in which machines can go beyond pattern recognition and start making inferences from data. The third stage will be reached only when we can create virtual human beings, who can think, act, and behave as humans do.

The dawn of artificial intelligence

The Economist, May 9, 2017.
https://www.economist.com/news/leaders/21650543-powerful-computers-will-reshape-humanitys-future-how-ensure-promise-outweighs

Editors' Note:
Here The Economist lays out a good understanding of the current state of AI technology and what possibilities lay in the future.
~

The first step is to understand what computers can now do and what they are likely to be able to do in the future. Thanks to the rise in processing power and the growing abundance of digitally available data, AI is enjoying a boom in its capabilities […]. Today's "deep learning" systems, by mimicking the layers of neurons in a human brain and crunching vast amounts of data, can teach themselves to perform some tasks, from pattern recognition to translation, almost as well as humans can. As a result, things that once called for a mind—from interpreting pictures to playing the video game "Frogger"—are now within the scope of computer programs. DeepFace, an algorithm unveiled by Facebook in 2014, can recognize individual human faces in images 97% of the time.

Crucially, this capacity is narrow and specific. Today's AI produces the semblance of intelligence through brute number-crunching force, without any great interest in approximating how minds equip humans with autonomy, interests and desires. Computers do not yet have anything approaching the wide, fluid ability to infer, judge and decide that is associated with intelligence in the conventional human sense.

Difference between automation and autonomous system

M.L Cummings, *Chatham House,* January 2017.
https://www.chathamhouse.org/sites/files/chathamhouse/publications/research/2017-01-26-artificial-intelligence-future-warfare-cummings-final.pdf

Editors' Note:
AI systems are often described as being able to act autonomously to some degree. How does that differ from a system that is purely automated? How this difference is reflected in the programming of computer systems is described below.
~

To better understand the nuances of AI, it is important first to understand the difference between an automated and an autonomous system. An automated system is one in which a computer reasons by a clear if–then–else, rule-based structure, and does so deterministically, meaning that for each input the system output will always be the same (except if something fails). An autonomous system is one that reasons probabilistically given a set of inputs, meaning that it makes guesses about best possible courses of action given sensor data input. Unlike automated systems, when given the same input autonomous systems will not necessarily produce the exact same behavior every time; rather, such systems will produce a range of behaviors.

Human intelligence generally follows a sequence known as the perception–cognition–action information processing loop, in that individuals perceive something in the world around them, think about what to do, and then, once they have weighed up the options, make a decision to act. AI is programmed to do something similar, in that a computer senses the world around it, and then processes the incoming information through optimization and verification algorithms, with a choice of action made in a fashion similar to that of humans.

While there are many parallels between human intelligence and AI, there are stark differences too. Every autonomous system that interacts in a dynamic environment must construct a world model and continually update that model. This means that the world must be perceived (or sensed through cameras, microphones and/or tactile sensors) and then reconstructed in such a way that the computer 'brain' has an effective and updated model of the world it is in before it can make decisions.

The fidelity of the world model and the timeliness of its updates are the keys to an effective autonomous system. Autonomous UAV navigation, for example, is relatively straightforward, since the world model according to which it operates consists simply of maps that indicate preferred routes, height obstacles and no-fly zones. Radars augment this model in real time by indicating which altitudes are clear of obstacles. GPS coordinates convey to the UAV where it needs to go, with the overarching goal of the GPS coordinate plan being not to take the aircraft into a no-fly zone or cause it to collide with an obstacle.

In comparison, navigation for driverless cars is much more difficult. Cars not only need similar mapping abilities, but they must also understand where all nearby vehicles, pedestrians and cyclists are, and where all these are going in the next few seconds. Driverless cars (and some drones) do this through a combination of sensors like LIDAR (Light Detection And Ranging), traditional radars, and stereoscopic computer vision. Thus the world model of a driverless car is much more advanced than that of a typical UAV, reflecting the complexity of the operating environment. A driverless car computer is required to track all the dynamics of all nearby vehicles and obstacles, constantly compute all possible points of intersection, and then estimate how it thinks traffic is going to behave in order to make a decision to act.

Indeed, this form of estimating or guessing what other drivers will do is a key component of how humans drive, but humans do this with little cognitive effort. It takes a computer significant computation power to keep track of all these variables while also trying to maintain and update its current world model. Given this immense problem of computation, in order to maintain safe execution times for action a driverless car will make best guesses based on probabilistic distributions. In effect, therefore, the car is guessing which path or action is best, given some sort of confidence interval.

INTRODUCTION

Rise of the machines

The Economist, May 9, 2015.
https://www.economist.com/news/briefing/21650526-artificial-intelligence-scares-peopleexcessively-so-rise-machines

Editors' Note:
This 2015 cover story from *The Economist*, a British publication, introduces the key issues surrounding artificial intelligence. It describes:

- the "machine learning" technology behind AI
- the narrow applications of existing AI systems (and those in development)
- the economic implications of AI
- the race between internet and technology companies to develop advanced AI systems
- the debate between AI-optimists, like Mark Zuckerberg, and AI-pessimists, like Elon Musk

~

ELON MUSK busies himself building other people's futures. A serial entrepreneur who made his first fortune in the early days of the world wide web, he has since helped found a solar-power company to generate green electricity, an electric-car firm to liberate motorists from the internal-combustion engine, and a rocketry business—SpaceX—to pursue his desire to see a human colony on Mars within his lifetime. It makes him the sort of technologist you would expect might look on tomorrow with unbridled optimism.

Not all future technology meets with his approval, though. In a speech in October at the Massachusetts Institute of Technology, Mr Musk described artificial intelligence (AI) as "summoning the demon", and the creation of a rival to human intelligence as probably the biggest threat facing the world. He is not alone. Nick Bostrom, a philosopher at the University of Oxford who helped develop the notion of "existential risks"—those that threaten humanity in general—counts advanced artificial intelligence as one such, alongside giant asteroid strikes and all-out nuclear war. Lord Rees, who used to run the Royal Society, Britain's foremost scientific body, has since founded the Centre for the Study of Existential Risk, in Cambridge, which takes the risks posed by AI just as seriously.

Such worries are a mirror image of the optimism suffusing the field itself, which has enjoyed rapid progress over the past couple of years. Firms such as Google, Facebook, Amazon and Baidu have got into an AI arms race, poaching researchers, setting up laboratories and buying start-ups. The insiders are not, by and large, fretting about being surpassed by their creations. Their business is not so much making new sorts of minds as it is removing some of the need for the old sort, by taking tasks that used to be things which only people could do and making them amenable to machines.

The torrent of data thrown off by the world's internet-connected computers, tablets and smartphones, and the huge amounts of computing power now available for processing that torrent, means that their algorithms are more and more capable of understanding languages, recognising images and the like. Business is taking notice. So are those who worry about technology taking away people's jobs. Lots of work depends on recognising patterns and translating symbols. If computers replace some of the people now doing this, either by providing an automated alternative or by making a few such workers far more productive, there will be more white collars in the dole queue.

Signs of the AI boom are everywhere. Last year, Google was rumoured to have paid $400m for DeepMind, a London-based AI startup. It snatched the firm from under the nose of Facebook, which boasts its own dedicated AI research laboratory, headed by Yann LeCun, a star researcher hired from New York University. Google once employed Andrew Ng, an AI guru from Stanford University—until Baidu poached him last year to head up a new, Silicon Valley-based lab of its own. Firms such as Narrative Science, in Chicago, which hopes to automate the writing of reports (and which is already used by Forbes, a business magazine, to cover basic financial stories), and Kensho, of Cambridge, Massachusetts, which aims to automate some of the work done by "quants" in the financial industry, have been showered in cash by investors. On April 13th IBM announced plans to use a version of its Watson computer—which crushed two puny human champions at an obscurantist American quiz show called Jeopardy! in 2011—to analyse health records, looking for medical insights.

Deep thought

Research into artificial intelligence is as old as computers themselves. Much of the current excitement concerns a subfield of it called "deep learning", a modern refinement of "machine learning", in which computers teach themselves tasks by crunching large sets of data. Algorithms created in this manner are a way of bridging a gap that bedevils all AI research: by and large, tasks that are hard for humans are easy for computers, and vice versa. The simplest computer can run rings around the brightest person when it comes to wading through complicated mathematical equations. At the same time, the most powerful computers have, in the past, struggled with things that people find trivial, such as recognising faces, decoding speech and identifying objects in images.

One way of understanding this is that for humans to do things they find difficult, such as solving differential equations, they have to write a set of formal rules. Turning those rules into a program is then pretty simple. For stuff human beings find easy, though, there is no similar need for explicit rules—and trying to create them can be hard. To take one famous example, both adults and children can easily distinguish a cat from a non-cat. But describing how they do so is almost impossible.

Machine learning is a way of getting computers to know things when they see them by producing for themselves the rules their programmers cannot specify. The machines do this with heavy-duty statistical analysis of lots and lots of data.

Many systems use an old and venerable piece of AI technology, the neural network, to develop the statistics that they need. Neural networks were invented in the 1950s by researchers who had the idea that, though they did not know what intelligence was, they did know that brains had it. And brains do their information processing not with transistors, but with neurons. If you could simulate those neurons—spindly, highly interlinked cells that pass electrochemical signals between themselves—then perhaps some sort of intelligent behaviour might emerge.

Caught by the net

Neurons are immensely complex. Even today, the simulations used in AI are a stick-figure cartoon of the real thing. But early results suggested that even the crudest networks might be good for some tasks. Chris Bishop, an AI researcher with Microsoft, points out that telephone companies have, since the 1960s, been using echo-cancelling algorithms discovered by neural networks. But after such early successes the idea lost its allure. The computing power then available limited the size of the networks that could be simulated, and this limited the technology's scope.

In the past few years, however, the remarkable number-crunching power of chips developed for the demanding job of drawing video-game graphics has revived interest. Early neural networks were limited to dozens or hundreds of neurons, usually organised as a single layer. The latest, used by the likes of Google, can simulate billions. With that many ersatz neurons available, researchers can afford to take another cue from the brain and organise them in distinct, hierarchical layers (see diagram). It is this use of interlinked layers that puts the "deep" into deep learning.

Each layer of the network deals with a different level of abstraction. To process an image, for example, the lowest layer is fed the raw images. It notes things like the brightness and colours of individual pixels, and how those properties are distributed across the image. The next layer combines these observations into more abstract categories, identifying edges, shadows and the like. The layer after that will analyse those edges and shadows in turn, looking for combinations that signify features such as eyes, lips and ears. And these can then be combined into a representation of a face—and indeed not just any face, but even a new image of a particular face that the network has seen before.

To make such networks useful, they must first be trained. For the machine to program itself for facial recognition, for instance, it will be presented with a "training set" of thousands of images. Some will contain faces and some will not. Each will be labelled as such by a human. The images act as inputs to the system; the labels ("face" or "not face") as outputs. The computer's task is to come up with a statistical rule that correlates inputs with the correct outputs. To do that, it will hunt at every level of abstraction for whatever features are common to those images showing faces. Once these correlations are good enough, the machine will be able, reliably, to tell faces from not-faces in its training set. The next step is to let it loose on a fresh set of images, to see if the facial-recognition rules it has extracted hold up in the real world.

By working from the bottom up in this way, machine-learning algorithms learn to recognise features, concepts and categories that humans understand but struggle to define in code. But such

algorithms were, for a long time, narrowly specialised. Programs often needed hints from their designers, in the form of hand-crafted bits of code that were specific to the task at hand—one set of tweaks for processing images, say, and another for voice recognition.

Earlier neural networks, moreover, had only a limited appetite for data. Beyond a certain point, feeding them more information did not boost their performance. Modern systems need far less hand-holding and tweaking. They can also make good use of as many data as you are able throw at them. And because of the internet, there are plenty of data to throw.

Big internet companies like Baidu, Google and Facebook sit on huge quantities of information generated by their users. Reams of e-mails; vast piles of search and buying histories; endless images of faces, cars, cats and almost everything else in the world pile up in their servers. The people who run those firms know that these data contain useful patterns, but the sheer quantity of information is daunting. It is not daunting for machines, though. The problem of information overload turns out to contain its own solution, especially since many of the data come helpfully pre-labelled by the people who created them. Fortified with the right algorithms, computers can use such annotated data to teach themselves to spot useful patterns, rules and categories within.

The results are impressive. In 2014 Facebook unveiled an algorithm called DeepFace that can recognise specific human faces in images around 97% of the time, even when those faces are partly hidden or poorly lit. That is on a par with what people can do. Microsoft likes to boast that the object-recognition software it is developing for Cortana, a digital personal assistant, can tell its users the difference between a picture of a Pembroke Welsh Corgi and a Cardigan Welsh Corgi, two dog breeds that look almost identical (see pictures). Some countries, including Britain, already use face-recognition technology for border control. And a system capable of recognising individuals from video footage has obvious appeal for policemen and spies. A report published on May 5th showed how America's spies use voice-recognition software to convert phone calls into text, in order to make their contents easier to search.

But, although the internet is a vast data trove, it is not a bottomless one. The sorts of human-labelled data that machine-learning algorithms thrive on are a finite resource. For this reason, a race is on to develop "unsupervised-learning" algorithms, which can learn without the need for human help.

There has already been lots of progress. In 2012 a team at Google led by Dr Ng showed an unsupervised-learning machine millions of YouTube video images. The machine learned to categorise common things it saw, including human faces and (to the amusement of the internet's denizens) the cats—sleeping, jumping or skateboarding—that are ubiquitous online. No human being had tagged the videos as containing "faces" or "cats". Instead, after seeing zillions of examples of each, the machine had simply decided that the statistical patterns they represented were common enough to make into a category of object.

The next step up from recognising individual objects is to recognise lots of different ones. A paper published by Andrej Karpathy and Li Fei-Fei at Stanford University describes a computer-vision system that is able to label specific parts of a given picture. Show it a breakfast table, for

instance, and it will identify the fork, the banana slices, the cup of coffee, the flowers on the table and the table itself. It will even generate descriptions, in natural English, of the scene (see picture right)—though the technology is not yet perfect (see picture below).

Big internet firms such as Google are interested in this kind of work because it can directly affect their bottom lines. Better image classifiers should improve the ability of search engines to find what their users are looking for. In the longer run, the technology could find other, more transformative uses. Being able to break down and interpret a scene would be useful for robotics researchers, for instance, helping their creations—from industrial helpmeets to self-driving cars to battlefield robots—to navigate the cluttered real world.

Image classification is also an enabling technology for "augmented reality", in which wearable computers, such as Google's Glass or Microsoft's HoloLens, overlay useful information on top of the real world. Enlitic, a firm based in San Francisco, hopes to employ image recognition to analyse X-rays and MRI scans, looking for problems that human doctors might miss.
And deep learning is not restricted to images. It is a general-purpose pattern-recognition technique, which means, in principle, that any activity which has access to large amounts of data—from running an insurance business to research into genetics—might find it useful. At a recent competition held at CERN, the world's biggest particle-physics laboratory, deep-learning algorithms did a better job of spotting the signatures of subatomic particles than the software written by physicists—even though the programmers who created these algorithms had no particular knowledge of physics. More whimsically, a group of researchers have written a program that learnt to play video games such as "Space Invaders" better than people can.

Machine translation, too, will be improved by deep learning. It already uses neural networks, benefiting from the large quantity of text available online in multiple languages. Dr Ng, now at Baidu, thinks good speech-recognition programs running on smartphones could bring the internet to many people in China who are illiterate, and thus struggle with ordinary computers. At the moment, 10% of the firm's searches are conducted by voice. He believes that could rise to 50% by 2020.

And those different sorts of AI can be linked together to form an even more capable system. In May 2014, for instance, at a conference in California, Microsoft demonstrated a computer program capable of real-time translation of spoken language. The firm had one of its researchers speak, in English, to a colleague in Germany. This colleague heard her interlocutor speaking in German. One AI program decoded sound waves into English phrases. Another translated those phrases from English into German, and a third rendered them into German speech. The firm hopes, one day, to build the technology into Skype, its internet-telephony service.

No ghost in the machine
Better smartphones, fancier robots and bringing the internet to the illiterate would all be good things. But do they justify the existential worries of Mr Musk and others? Might pattern-recognising, self-programming computers be an early, but crucial, step on the road to machines that are more intelligent than their creators?

The doom-mongers have one important fact on their side. There is no result from decades of neuroscientific research to suggest that the brain is anything other than a machine, made of ordinary atoms, employing ordinary forces and obeying the ordinary laws of nature. There is no mysterious "vital spark", in other words, that is necessary to make it go. This suggests that building an artificial brain—or even a machine that looks different from a brain but does the same sort of thing—is possible in principle.

But doing something in principle and doing it in fact are not remotely the same thing. Part of the problem, says Rodney Brooks, who was one of AI's pioneers and who now works at Rethink Robotics, a firm in Boston, is a confusion around the word "intelligence". Computers can now do some narrowly defined tasks which only human brains could manage in the past (the original "computers", after all, were humans, usually women, employed to do the sort of tricky arithmetic that the digital sort find trivially easy). An image classifier may be spookily accurate, but it has no goals, no motivations, and is no more conscious of its own existence than is a spreadsheet or a climate model. Nor, if you were trying to recreate a brain's workings, would you necessarily start by doing the things AI does at the moment in the way that it now does them. AI uses a lot of brute force to get intelligent-seeming responses from systems that, though bigger and more powerful now than before, are no more like minds than they ever were. It does not seek to build systems that resemble biological minds. As Edsger Dijkstra, another pioneer of AI, once remarked, asking whether a computer can think is a bit like asking "whether submarines can swim".

Nothing makes this clearer than the ways in which AI programs can be spoofed. A paper to be presented at a computer-vision conference in June shows optical illusions designed to fool image-recognition algorithms (see picture). These offer insight into how the algorithms operate—by matching patterns to other patterns, but doing so blindly, with no recourse to the sort of context (like realising a baseball is a physical object, not just an abstract pattern vaguely reminiscent of stitching) that stops people falling into the same traps. It is even possible to construct images that, to a human, look like meaningless television static, but which neural networks nevertheless confidently classify as real objects.

This is not to say that progress in AI will have no unpleasant consequences, at least for some people. And, unlike previous waves of technological change, quite a few of those people may be middle class. Take Microsoft's real-time translation. The technology it demonstrated was far from perfect. No one would mistake its computer-translated speech for the professionally translated sort. But it is adequate to convey the gist of what is being said. It is also cheaper and more convenient than hiring a human interpreter. Such an algorithm could therefore make a limited version of what is presently a costly, bespoke service available to anyone with a Skype account. That might be bad for interpreters. But it would be a boon for everyone else. And Microsoft's program will only get better.

The worry that AI could do to white-collar jobs what steam power did to blue-collar ones during the Industrial Revolution is therefore worth taking seriously. Examples, such as Narrative Science's digital financial journalist and Kensho's quant, abound. Kensho's system is designed to interpret natural-language search queries such as, "What happens to car firms' share prices if

oil drops by $5 a barrel?" It will then scour financial reports, company filings, historical market data and the like, and return replies, also in natural language, in seconds. The firm plans to offer the software to big banks and sophisticated traders. Yseop, a French firm, uses its natural-language software to interpret queries, chug through data looking for answers, and then write them up in English, Spanish, French or German at 3,000 pages a second. Firms such as L'Oréal and VetOnline.com already use it for customer support on their websites.

Nor is this just a theoretical worry, for some white-collar jobs are already being lost to machines. Many firms use computers to answer telephones, for instance. For all their maddening limitations, and the need for human backup when they encounter a query they cannot understand, they are cheaper than human beings. Forecasting how many more jobs might go the same way is much harder—although a paper from the Oxford Martin School, published in 2013, scared plenty of people by concluding that up to half of the job categories tracked by American statisticians might be vulnerable.

Technology, though, gives as well as taking away. Automated, cheap translation is surely useful. Having an untiring, lightning-fast computer checking medical images would be as well. Perhaps the best way to think about AI is to see it as simply the latest in a long line of cognitive enhancements that humans have invented to augment the abilities of their brains. It is a high-tech relative of technologies like paper, which provides a portable, reliable memory, or the abacus, which aids mental arithmetic. Just as the printing press put scribes out of business, high-quality AI will cost jobs. But it will enhance the abilities of those whose jobs it does not replace, giving everyone access to mental skills possessed at present by only a few. These days, anyone with a smartphone has the equivalent of a city-full of old-style human "computers" in his pocket, all of them working for nothing more than the cost of charging the battery. In the future, they might have translators or diagnosticians at their beck and call as well.

Cleverer computers, then, could be a truly transformative technology, though not—at least, not yet—for the reasons given by Mr Musk or Lord Rees. One day, perhaps, something like the sort of broad intelligence that characterizes the human brain may be recreated in a machine. But for now, the best advice is to ignore the threat of computers taking over the world—and check that they are not going to take over your job first.

The Great A.I. Awakening: How Google used artificial intelligence to transform Google Translate, one of its more popular services — and how machine learning is poised to reinvent computing itself.

Gideon Lewis-Kraus, *The New York Times Magazine*, December 14, 2016.
https://www.nytimes.com/2016/12/14/magazine/the-great-ai-awakening.html

Editors' Note:
This long-form article traces the history of Google Brain, which was founded in 2011 to conduct research into artificial intelligence. Over a 9-month period in 2016, Google overhauled Google Translate with machine learning" software developed by Brain, and the accuracy of Translate dramatically improved.

The history of Google Brain illustrates several critical points about AI. You do not need to know the names of the particular scientists involved, so what follows are edited excerpts of the article that capture the most important points.

First, the author explains why technology giants like Google are reorganizing their business models around the research and development of AI. Piecemeal AI innovations—like translation software—not only improve the existing services of tech companies. They also lay the groundwork for the development of "artificial general intelligence."

Second, the article uses image-recognition and language-recognition software as examples to explain how AI works—and how it is distinct from traditional computer programming. As the preceding article from *The Economist* noted, traditional computer programs are poor at classifying images and understanding language. But AI systems, which mimic how the human brain processes information, have proven very successful at them. We recommend that debaters read the "Machine Learning", "Artificial Neural Networks", and "Language Recognition" sections to gain a better understanding of the technology behind AI.

Finally, throughout the piece, the author reflects on questions that will be explored in the "Applications" and "Issues" section of this research packet. What are the economic implications of AI? Will the proliferation of AI-based systems in society, government, and industry make tech companies too powerful? Are AI systems controllable? What are their shortcomings?
~

Introduction
Google Translate made its debut in 2006 and since then has become one of Google's most reliable and popular assets; it serves more than 500 million monthly users in need of 140 billion words per day in a different language. It exists not only as its own stand-alone app but also as an integrated feature within Gmail, Chrome and many other Google offerings, where we take it as a push-button given — a frictionless, natural part of our digital commerce. The Translate team had

been steadily adding new languages and features, but gains in quality over the last four years had slowed considerably.

Until today. As of the previous weekend, Translate had been converted to an A.I.-based system for much of its traffic, not just in the United States but in Europe and Asia as well: The rollout included translations between English and Spanish, French, Portuguese, German, Chinese, Japanese, Korean and Turkish. The rest of Translate's hundred-odd languages were to come, with the aim of eight per month, by the end of next year. The new incarnation, to the pleasant surprise of Google's own engineers, had been completed in only nine months. The A.I. system had demonstrated overnight improvements roughly equal to the total gains the old one had accrued over its entire lifetime.

[…]

Google's decision to reorganize itself around A.I. was the first major manifestation of what has become an industrywide machine-learning delirium. Over the past four years, six companies in particular — Google, Facebook, Apple, Amazon, Microsoft and the Chinese firm Baidu — have touched off an arms race for A.I. talent, particularly within universities. Corporate promises of resources and freedom have thinned out top academic departments. It has become widely known in Silicon Valley that Mark Zuckerberg, chief executive of Facebook, personally oversees, with phone calls and video-chat blandishments, his company's overtures to the most desirable graduate students. Starting salaries of seven figures are not unheard-of. Attendance at the field's most important academic conference has nearly quadrupled. What is at stake is not just one more piecemeal innovation but control over what very well could represent an entirely new computational platform: pervasive, ambient artificial intelligence.

The phrase "artificial intelligence" is invoked as if its meaning were self-evident, but it has always been a source of confusion and controversy. Imagine if you went back to the 1970s, stopped someone on the street, pulled out a smartphone and showed her Google Maps. Once you managed to convince her you weren't some oddly dressed wizard, and that what you withdrew from your pocket wasn't a black-arts amulet but merely a tiny computer more powerful than the one that guided Apollo missions, Google Maps would almost certainly seem to her a persuasive example of "artificial intelligence." In a very real sense, it is. It can do things any map-literate human can manage, like get you from your hotel to the airport — though it can do so much more quickly and reliably. It can also do things that humans simply and obviously cannot: It can evaluate the traffic, plan the best route and reorient itself when you take the wrong exit.

Practically nobody today, however, would bestow upon Google Maps the honorific "A.I.," so sentimental and sparing are we in our use of the word "intelligence." The minute we can automate a task, we downgrade the relevant skill involved to one of mere mechanism. Today Google Maps seems, in the pejorative sense of the term, robotic: It simply accepts an explicit demand (the need to get from one place to another) and tries to satisfy that demand as efficiently as possible. The goal posts for "artificial intelligence" are thus constantly receding.

When he has an opportunity to make careful distinctions, Sundar Pichai [the chief executive of Google] differentiates between the current applications of A.I. and the ultimate goal of "artificial general intelligence." Artificial general intelligence will not involve dutiful adherence to explicit instructions, but instead will demonstrate a facility with the implicit, the interpretive. It will be a general tool, designed for general purposes in a general context. Pichai believes his company's future depends on something like this. Imagine if you could tell Google Maps, "I'd like to go to the airport, but I need to stop off on the way to buy a present for my nephew." A more generally intelligent version of that service — a ubiquitous assistant, of the sort that Scarlett Johansson memorably disembodied three years ago in the Spike Jonze film "Her"— would know all sorts of things that, say, a close friend or an earnest intern might know: your nephew's age, and how much you ordinarily like to spend on gifts for children, and where to find an open store. But a truly intelligent Maps could also conceivably know all sorts of things a close friend wouldn't, like what has only recently come into fashion among preschoolers in your nephew's school — or more important, what its users actually want. If an intelligent machine were able to discern some intricate if murky regularity in data about what we have done in the past, it might be able to extrapolate about our subsequent desires, even if we don't entirely know them ourselves.

The new wave of A.I.-enhanced assistants — Apple's Siri, Facebook's M, Amazon's Echo — are all creatures of machine learning, built with similar intentions. The corporate dreams for machine learning, however, aren't exhausted by the goal of consumer clairvoyance. A medical-imaging subsidiary of Samsung announced this year that its new ultrasound devices could detect breast cancer. Management consultants are falling all over themselves to prep executives for the widening industrial applications of computers that program themselves. DeepMind, a 2014 Google acquisition, defeated the reigning human grandmaster of the ancient board game Go, despite predictions that such an achievement would take another 10 years.

In a famous 1950 essay, Alan Turing proposed a test for an artificial general intelligence: a computer that could, over the course of five minutes of text exchange, successfully deceive a real human interlocutor. Once a machine can translate fluently between two natural languages, the foundation has been laid for a machine that might one day "understand" human language well enough to engage in plausible conversation. Google Brain's members, who pushed and helped oversee the Translate project, believe that such a machine would be on its way to serving as a generally intelligent all-encompassing personal digital assistant. The overhaul of Google Translate made considerable progress in that direction.

[…]

Machine Learning
Since the term "artificial intelligence" was first coined, at a kind of constitutional convention of the mind at Dartmouth in the summer of 1956, a majority of researchers have long thought the best approach to creating A.I. would be to write a very big, comprehensive program that laid out both the rules of logical reasoning and sufficient knowledge of the world. If you wanted to translate from English to Japanese, for example, you would program into the computer all of the grammatical rules of English, and then the entirety of definitions contained in the Oxford English

Dictionary, and then all of the grammatical rules of Japanese, as well as all of the words in the Japanese dictionary, and only after all of that feed it a sentence in a source language and ask it to tabulate a corresponding sentence in the target language. You would give the machine a language map that was, as Borges would have had it, the size of the territory. This perspective is usually called "symbolic A.I." — because its definition of cognition is based on symbolic logic — or, disparagingly, "good old-fashioned A.I."

There are two main problems with the old-fashioned approach. The first is that it's awfully time-consuming on the human end. The second is that it only really works in domains where rules and definitions are very clear: in mathematics, for example, or chess. Translation, however, is an example of a field where this approach fails horribly, because words cannot be reduced to their dictionary definitions, and because languages tend to have as many exceptions as they have rules. More often than not, a system like this is liable to translate "minister of agriculture" as "priest of farming." Still, for math and chess it worked great, and the proponents of symbolic A.I. took it for granted that no activities signaled "general intelligence" better than math and chess.

There were, however, limits to what this system could do. In the 1980s, a robotics researcher at Carnegie Mellon pointed out that it was easy to get computers to do adult things but nearly impossible to get them to do things a 1-year-old could do, like hold a ball or identify a cat. By the 1990s, despite punishing advancements in computer chess, we still weren't remotely close to artificial general intelligence.

There has always been another vision for A.I. — a dissenting view — in which the computers would learn from the ground up (from data) rather than from the top down (from rules). This notion dates to the early 1940s, when it occurred to researchers that the best model for flexible automated intelligence was the brain itself. A brain, after all, is just a bunch of widgets, called neurons, that either pass along an electrical charge to their neighbors or don't. What's important are less the individual neurons themselves than the manifold connections among them. This structure, in its simplicity, has afforded the brain a wealth of adaptive advantages. The brain can operate in circumstances in which information is poor or missing; it can withstand significant damage without total loss of control; it can store a huge amount of knowledge in a very efficient way; it can isolate distinct patterns but retain the messiness necessary to handle ambiguity.

There was no reason you couldn't try to mimic this structure in electronic form, and in 1943 it was shown that arrangements of simple artificial neurons could carry out basic logical functions. They could also, at least in theory, learn the way we do. With life experience, depending on a particular person's trials and errors, the synaptic connections among pairs of neurons get stronger or weaker. An artificial neural network could do something similar, by gradually altering, on a guided trial-and-error basis, the numerical relationships among artificial neurons. It wouldn't need to be preprogrammed with fixed rules. It would, instead, rewire itself to reflect patterns in the data it absorbed.

This attitude toward artificial intelligence was evolutionary rather than creationist. If you wanted a flexible mechanism, you wanted one that could adapt to its environment. If you wanted

something that could adapt, you didn't want to begin with the indoctrination of the rules of chess. You wanted to begin with very basic abilities — sensory perception and motor control — in the hope that advanced skills would emerge organically. Humans don't learn to understand language by memorizing dictionaries and grammar books, so why should we possibly expect our computers to do so? Google Brain was the first major commercial institution to invest in the possibilities embodied by this way of thinking about A.I. Its speech-recognition team swapped out part of their old system for a neural network and encountered, in pretty much one fell swoop, the best quality improvements anyone had seen in 20 years.

[…]

Artificial Neural Networks

An average brain has something on the order of 100 billion neurons. Each neuron is connected to up to 10,000 other neurons, which means that the number of synapses is between 100 trillion and 1,000 trillion. For a simple artificial neural network of the sort proposed in the 1940s, the attempt to even try to replicate this was unimaginable. We're still far from the construction of a network of that size, but Google Brain's investment allowed for the creation of artificial neural networks comparable to the brains of mice.

To understand why scale is so important, however, you have to start to understand some of the more technical details of what, exactly, machine intelligences are doing with the data they consume. A lot of our ambient fears about A.I. rest on the idea that they're just vacuuming up knowledge like a sociopathic prodigy in a library, and that an artificial intelligence constructed to make paper clips might someday decide to treat humans like ants or lettuce. This just isn't how they work. All they're doing is shuffling information around in search of commonalities — basic patterns, at first, and then more complex ones — and for the moment, at least, the greatest danger is that the information we're feeding them is biased in the first place.

Imagine you want to program a computer to recognize cats on the old symbolic-A.I. model. You stay up for days preloading the machine with an exhaustive, explicit definition of "cat." You tell it that a cat has four legs and pointy ears and whiskers and a tail, and so on. All this information is stored in a special place in memory called Cat. Now you show it a picture. First, the machine has to separate out the various distinct elements of the image. Then it has to take these elements and apply the rules stored in its memory. If(legs=4) and if(ears=pointy) and if(whiskers=yes) and if(tail=yes) and if(expression=supercilious), then(cat=yes). But what if you showed this cat-recognizer a Scottish Fold, a heart-rending breed with a prized genetic defect that leads to droopy doubled-over ears? Our symbolic A.I. gets to (ears=pointy) and shakes its head solemnly, "Not cat." It is hyperliteral, or "brittle." Even the thickest toddler shows much greater inferential acuity.

Now imagine that instead of hard-wiring the machine with a set of rules for classification stored in one location of the computer's memory, you try the same thing on a neural network. There is no special place that can hold the definition of "cat." There is just a giant blob of interconnected switches, like forks in a path. On one side of the blob, you present the inputs (the pictures); on the other side, you present the corresponding outputs (the labels). Then you just tell it to work

out for itself, via the individual calibration of all of these interconnected switches, whatever path the data should take so that the inputs are mapped to the correct outputs. The training is the process by which a labyrinthine series of elaborate tunnels are excavated through the blob, tunnels that connect any given input to its proper output. The more training data you have, the greater the number and intricacy of the tunnels that can be dug. Once the training is complete, the middle of the blob has enough tunnels that it can make reliable predictions about how to handle data it has never seen before. This is called "supervised learning."

Part of the reason there was so much resistance to these ideas in computer-science departments is that because the output is just a prediction based on patterns of patterns, it's not going to be perfect, and the machine will never be able to define for you what, exactly, a cat is. It just knows them when it sees them.

The fact that neural networks are probabilistic in nature means that they're not suitable for all tasks. It's no great tragedy if they mislabel 1 percent of cats as dogs, or send you to the wrong movie on occasion, but in something like a self-driving car we all want greater assurances. This isn't the only caveat. Supervised learning is a trial-and-error process based on labeled data. The machines might be doing the learning, but there remains a strong human element in the initial categorization of the inputs. If your data had a picture of a man and a woman in suits that someone had labeled "woman with her boss," that relationship would be encoded into all future pattern recognition. Labeled data is thus fallible the way that human labelers are fallible. If a machine was asked to identify creditworthy candidates for loans, it might use data like felony convictions, but if felony convictions were unfair in the first place — if they were based on, say, discriminatory drug laws — then the loan recommendations would perforce also be fallible.

[…]

Language Recognition
If you took the entire space of the English language and the entire space of French, you could, at least in theory, train a neural network to learn how to take a sentence in one space and propose an equivalent in the other. You just had to give it millions and millions of English sentences as inputs on one side and their desired French outputs on the other, and over time it would recognize the relevant patterns in words the way that an image classifier recognized the relevant patterns in pixels. You could then give it a sentence in English and ask it to predict the best French analogue. This is the strategy used by Google Brain to transform Google's translation software.

The old system of Google Translate worked the way all machine translation has worked for about 30 years: It sequestered each successive sentence fragment, looked up those words in a large statistically derived vocabulary table, then applied a battery of post-processing rules to affix proper endings and rearrange it all to make sense. The approach is called "phrase-based statistical machine translation," because by the time the system gets to the next phrase, it doesn't know what the last one was. This is why Translate's output sometimes looked like a shaken bag of fridge magnets. Brain's replacement would, if it came together, read and render entire sentences at one draft. It would capture context — and something akin to meaning.

The stakes may have seemed low: Translate generates minimal revenue, and it probably always will. For most Anglophone users, even a radical upgrade in the service's performance would hardly be hailed as anything more than an expected incremental bump. But there was a case to be made that human-quality machine translation is not only a short-term necessity but also a development very likely, in the long term, to prove transformational. In the immediate future, it's vital to the company's business strategy. Google estimates that 50 percent of the internet is in English, which perhaps 20 percent of the world's population speaks. If Google was going to compete in China — where a majority of market share in search-engine traffic belonged to its competitor Baidu — or India, decent machine translation would be an indispensable part of the infrastructure. Baidu itself had published a pathbreaking paper about the possibility of neural machine translation in July 2015.

And in the more distant, speculative future, machine translation was perhaps the first step toward a general computational facility with human language. This would represent a major inflection point — perhaps the major inflection point — in the development of something that felt like true artificial intelligence.

By late spring, the various pieces were coming together at Google. The team introduced something called a "word-piece model," a "coverage penalty," "length normalization." Each part improved the results by maybe a few percentage points, but in aggregate they had significant effects. Once the model was standardized, it would be only a single multilingual model that would improve over time, rather than the 150 different models that Translate currently used.

By late September, after a mere nine months, Google Brain's neural translation was working.

[…]

Conclusion

Perhaps the most famous historic critique of artificial intelligence, or the claims made on its behalf, implicates the question of translation. The Chinese Room argument was proposed in 1980 by the Berkeley philosopher John Searle. In Searle's thought experiment, a monolingual English speaker sits alone in a cell. An unseen jailer passes him, through a slot in the door, slips of paper marked with Chinese characters. The prisoner has been given a set of tables and rules in English for the composition of replies. He becomes so adept with these instructions that his answers are soon "absolutely indistinguishable from those of Chinese speakers." Should the unlucky prisoner be said to "understand" Chinese? Searle thought the answer was obviously not. This metaphor for a computer, Searle later wrote, exploded the claim that "the appropriately programmed digital computer with the right inputs and outputs would thereby have a mind in exactly the sense that human beings have minds."

For the Google Brain team, though, or for nearly everyone else who works in machine learning in Silicon Valley, that view is entirely beside the point. This doesn't mean they're just ignoring the philosophical question. It means they have a fundamentally different view of the mind. Unlike Searle, they don't assume that "consciousness" is some special, numinously glowing

mental attribute — what the philosopher Gilbert Ryle called the "ghost in the machine." They just believe instead that the complex assortment of skills we call "consciousness" has randomly emerged from the coordinated activity of many different simple mechanisms. The implication is that our facility with what we consider the higher registers of thought are no different in kind from what we're tempted to perceive as the lower registers. Logical reasoning, on this account, is seen as a lucky adaptation; so is the ability to throw and catch a ball. Artificial intelligence is not about building a mind; it's about the improvement of tools to solve problems. As Corrado said to me on my very first day at Google, "It's not about what a machine 'knows' or 'understands' but what it 'does,' and — more importantly — what it doesn't do yet."

Where you come down on "knowing" versus "doing" has real cultural and social implications. At the party, Schuster came over to me to express his frustration with the paper's media reception. "Did you see the first press?" he asked me. He paraphrased a headline from that morning, blocking it word by word with his hand as he recited it: GOOGLE SAYS A.I. TRANSLATION IS INDISTINGUISHABLE FROM HUMANS'. Over the final weeks of the paper's composition, the team had struggled with this; Schuster often repeated that the message of the paper was "It's much better than it was before, but not as good as humans." He had hoped it would be clear that their efforts weren't about replacing people but helping them.

And yet the rise of machine learning makes it more difficult for us to carve out a special place for us. If you believe, with Searle, that there is something special about human "insight," you can draw a clear line that separates the human from the automated. If you agree with Searle's antagonists, you can't. It is understandable why so many people cling fast to the former view. At a 2015 M.I.T. conference about the roots of artificial intelligence, Noam Chomsky was asked what he thought of machine learning. He pooh-poohed the whole enterprise as mere statistical prediction, a glorified weather forecast. Even if neural translation attained perfect functionality, it would reveal nothing profound about the underlying nature of language. It could never tell you if a pronoun took the dative or the accusative case. This kind of prediction makes for a good tool to accomplish our ends, but it doesn't succeed by the standards of furthering our understanding of why things happen the way they do. A machine can already detect tumors in medical scans better than human radiologists, but the machine can't tell you what's causing the cancer. Then again, can the radiologist?

Medical diagnosis is one field most immediately, and perhaps unpredictably, threatened by machine learning. Radiologists are extensively trained and extremely well paid, and we think of their skill as one of professional insight — the highest register of thought. In the past year alone, researchers have shown not only that neural networks can find tumors in medical images much earlier than their human counterparts but also that machines can even make such diagnoses from the texts of pathology reports. What radiologists do turns out to be something much closer to predictive pattern-matching than logical analysis. They're not telling you what caused the cancer; they're just telling you it's there.

Once you've built a robust pattern-matching apparatus for one purpose, it can be tweaked in the service of others. One Translate engineer took a network he put together to judge artwork and used it to drive an autonomous radio-controlled car. A network built to recognize a cat can be

turned around and trained on CT scans — and on infinitely more examples than even the best doctor could ever review. A neural network built to translate could work through millions of pages of documents of legal discovery in the tiniest fraction of the time it would take the most expensively credentialed lawyer. The kinds of jobs taken by automatons will no longer be just repetitive tasks that were once — unfairly, it ought to be emphasized — associated with the supposed lower intelligence of the uneducated classes. We're not only talking about three and a half million truck drivers who may soon lack careers. We're talking about inventory managers, economists, financial advisers, real estate agents. What Brain did over nine months is just one example of how quickly a small group at a large company can automate a task nobody ever would have associated with machines.

The most important thing happening in Silicon Valley right now is not disruption. Rather, it's institution-building — and the consolidation of power — on a scale and at a pace that are both probably unprecedented in human history. Brain has interns; it has residents; it has "ninja" classes to train people in other departments. Everywhere there are bins of free bike helmets, and free green umbrellas for the two days a year it rains, and little fruit salads, and nap pods, and shared treadmill desks, and massage chairs, and random cartons of high-end pastries, and places for baby-clothes donations, and two-story climbing walls with scheduled instructors, and reading groups and policy talks and variegated support networks. The recipients of these major investments in human cultivation — for they're far more than perks for proles in some digital salt mine — have at hand the power of complexly coordinated servers distributed across 13 data centers on four continents, data centers that draw enough electricity to light up large cities.

But even enormous institutions like Google will be subject to this wave of automation; once machines can learn from human speech, even the comfortable job of the programmer is threatened.

APPLICATIONS OF AI

GENERAL

Eight ways intelligent machines are already in your life

Dr. Sabine Hauert, *BBC News*, April 25, 2017.
http://www.bbc.com/news/uk-39657382

Many people are unsure about exactly what machine learning is. But the reality is that it is already part of everyday life.

A form of artificial intelligence, it allows computers to learn from examples rather than having to follow step-by-step instructions.

The Royal Society believes it will have an increasing impact on people's lives and is calling for more research, to ensure the UK makes the most of opportunities.

Machine learning is already powering systems from the seemingly mundane to the life-changing. Here are just a few examples.

1. On your phone
Using spoken commands to ask your phone to carry out a search, or make a call, relies on technology supported by machine learning.

Virtual personal assistants - the likes of Siri, Alexa, Cortana and Google Assistant - are able to follow instructions because of voice recognition.

They process natural human speech, match it to the desired command and respond in an increasingly natural way.

The assistants learn over a number of conversations and in many different ways.

They might ask for specific information - for example how to pronounce your name, or whose voice is whose in a household.

Data from large numbers of conversations by all users is also sampled, to help them recognise words with different pronunciations or how to create natural discussion.

2. In your shopping basket
Many of us are familiar with shopping recommendations - think of the supermarket that reminds you to add cheese to your online shop, or the way Amazon suggests books it thinks you might like.

Machine learning is the technology that helps deliver these suggestions, via so-called recommender systems.

By analysing data about what customers have bought before, and any preferences they have expressed, recommender systems can pick up on patterns in purchasing history. They use this to make predictions about the products you might like.

3. On your TV
Similar systems are used to recommend films or TV shows on streaming services like Netflix. Recommender systems use machine learning to analyse viewing habits and pick out patterns in who watches - and enjoys - which shows.

By understanding which users like which films - and what shows you have watched or awarded high ratings - recommender systems can identify your tastes.

They are also used to suggest music on streaming services, like Spotify, and articles to read on Facebook.

4. In your email
Machine learning can also be used to distinguish between different categories of objects or items.

This makes it useful when sorting out the emails you want to see from those you don't.

Spam detection systems use a sample of emails to work out what is junk - learning to detect the presence of specific words, the names of certain senders, or other characteristics.

Once deployed, the system uses this learning to direct emails to the right folder. It continues to learn as users flag emails, or move them between folders.

5. On your social media
Ever wondered how Facebook knows who is in your photos and can automatically label your pictures?

The image recognition systems that Facebook - and other social media - uses to automatically tag photos is based on machine learning.

When users upload images and tag their friends and family, these image recognition systems can spot pictures that are repeated and assigns these to categories - or people.

6. At your bank
By analysing large amounts of data and looking for patterns, activity which might not otherwise be visible to human analysts can be identified.

One common application of this ability is in the fight against debit and credit card fraud.

Machine learning systems can be trained to recognise typical spending patterns and which characteristics of a transaction - location, amount, or timing - make it more or less likely to be fraudulent.

When a transaction seems out of the ordinary, an alarm can be raised - and a message sent to the user.

7. In hospitals
Doctors are just starting to consider machine learning to make better diagnoses, for example to spot cancer and eye disease.

Learning from images that have been labelled by doctors, computers can analyse new pictures of a patient's retina, a skin spot, or an image of cells taken under a microscope.

In doing so, they look for visual clues that indicate the presence of medical conditions.

This type of image recognition system is increasingly important in healthcare diagnostics.

8. In science
Machine learning is also powering scientists' ability to make new discoveries.

In particle physics it has allowed them to find patterns in immense data sets generated from the Large Hadron Collider at Cern.

It was instrumental in the discovery of the Higgs Boson, for example, and is now being used to search for "new physics" that no-one has yet imagined.

Similar ideas are being used to search for new medicines, for example by looking for new small molecules and antibodies to fight diseases.

What next?
The focus will be on making systems that perform specific tasks well which could therefore be thought of as helpers.

In schools they could track student performance and develop personal learning plans.

They could help us reduce energy usage by making better use of resources and improve care for the elderly by finding more time for meaningful human contact.

In the area of transport, machine learning will power autonomous vehicles.

Many industries could turn to algorithms to increase productivity. Financial services could become increasingly automated and law firms may use machine learning to carry out basic research.

Routine tasks will be done faster, challenging business models that rely on charging hourly rates.

Over the next 10 years machine learning technologies will increasingly be part of our lives, transforming the way we work and live.

BUSINESS

Meet the Company That's Using Face Recognition to Reshape China's Tech Scene

Yiting Sun, *MIT Technology Review*, August 11, 2017
https://www.technologyreview.com/s/608598/when-a-face-is-worth-a-billion-dollars/

Editors' Note:
For more analysis on how AI-based targeted products raise privacy questions, see the section on privacy concerns.
~

In China, face recognition is transforming many aspects of daily life. Employees at e-commerce giant Alibaba in Shenzhen can show their faces to enter their office building instead of swiping ID cards. A train station in western Beijing matches passengers' tickets to their government-issued IDs by scanning their faces. If their face matches their ID card photo, the system deems their tickets valid and the station gate will open. The subway system in Hangzhou, a city about 125 miles southwest of Shanghai, employs surveillance cameras capable of recognizing faces to spot suspected criminals.

The technology powering many of these applications? Face++, the world's largest face-recognition technology platform, currently used by more than 300,000 developers in 150 countries to identify faces, as well as images, text, and various kinds of government-issued IDs.

Other Chinese companies, such as Baidu and the startup SenseTime, also provide face-recognition technology to developers, but Face++'s popularity has been a boon for Megvii, the Beijing-based company that created and runs the platform. Founded in 2011 by three Tsinghua University graduates, Megvii is now valued at roughly a billion dollars and boasts approximately 530 employees, up from about 30 employees in 2014.

Megvii believes that as the Internet takes over more and more commercial and social functions, face recognition will become part of the Web's infrastructure as a means of identification, though only for activities that require real identities. Other tech companies seem to be betting on this scenario, too; Samsung's Galaxy S8 and S8+ phones support face recognition (for unlocking the devices), and Apple is rumored to be equipping its upcoming iPhone 8 with the technology.

Face ID, Megvii's online identity authentication platform, is one way Face++ is being integrated into the Internet's infrastructure. (Face ID's face-comparison API interface utilizes Face++ technology.) Nearly 90 percent of China's roughly 200 top Internet companies use Face ID, according to Yin. It's particularly popular with online financial services since they need to authenticate user identities remotely. (To avoid people tricking them with a photograph, these apps usually perform a "liveness test" that requires users to speak or move their heads.)

Xiaohua, which operates a virtual bank that grants loans and offers payments by installments through a mobile app called Xiaohua Qianbao ("Little Flower Wallet"), is a typical Face ID customer. Users scan their face using the app to get approved for loans and to ensure that nobody can authorize actions in the app if their phone is lost or stolen. "Xiaohua Qianbao is a purely online borrowing and lending product, so our first need is fraud prevention," says Lingpeng Huang, a cofounder of Xiaohua. "Face recognition has eliminated the risk of fake identities."

Megvii trains the algorithms that power Face++ and Face ID by feeding large data sets into a deep-learning engine called Brain++. (Deep learning involves feeding examples into a large, many-layered neural network, and tweaking its parameters until it accurately recognizes the desired features, such as a particular person's face.)

To amass huge amounts of training data, Megvii let most developers use Face++ for free during the platform's first two years of availability in 2012 and 2013. Megvii also purchases photos from data-collection companies to aid its training.

The company built Brain++ in 2015 and says having a self-developed deep-learning engine helps it train its algorithms more efficiently. "[It] translates into more competitiveness for our products," says Jian Sun, Megvii's chief scientist.

Another advantage of having its own deep-learning platform: Megvii can customize its face-recognition technology for different customers easily. That matters because a police department, for example, will value accuracy above everything else, but a company looking to use face recognition in a mobile app needs to ensure that the software is small enough to fit inside the app—without sacrificing too much accuracy.

When CEO Qi Yin launched Megvii, he wanted to gain traction in a few key areas. "An AI company has to be No. 1 in one or two core industries first [to succeed]," he says.

Now that Face++ is entrenched in banking and finance, Megvii's cofounders have plans to integrate its face recognition and other computer-vision technologies into more industries, such as retail and self-driving cars. For that to happen, the company needs to show those industries how they will benefit, such as how much fraud its technology can reduce every year, says Jiansheng Chen, an associate professor at Tsinghua University who studies computer vision.

LEGAL SYSTEM

Rise of the Robolawyers

Jason Koebler, *The Atlantic*, April 2017.
https://www.theatlantic.com/magazine/archive/2017/04/rise-of-the-robolawyers/517794/

Editors' Note:
This article explains how Artificial Intelligence will transform the legal industry and in turn the judicial system in the United States. These changes could have positive effects such as streamlining inefficient legal processes and increasing access to legal aid. However, there might be an increase in trivial lawsuits and an entrenchment of current racial biases.

The section on privacy concerns is also relevant to the legal system. For more analysis on the regulatory challenges posed by AI, see the sections on health care, the military, transportation, and macroeconomics.
~

EXPLANATION

Near the end of Shakespeare's Henry VI, Part 2, Dick the Butcher offers a simple plan to create chaos and help his band of outsiders ascend to the throne: "Let's kill all the lawyers." Though far from the Bard's most beautiful turn of phrase, it is nonetheless one of his most enduring. All these years later, the law is still America's most hated profession and one of the least trusted, whether you go by scientific studies or informal opinion polls.

Thankfully, no one's out there systematically murdering lawyers. But advances in artificial intelligence may diminish their role in the legal system or even, in some cases, replace them altogether. Here's what we stand to gain—and what we should fear—from these technologies.

1 | Handicapping Lawsuits
For years, artificial intelligence has been automating tasks—like combing through mountains of legal documents and highlighting keywords—that were once rites of passage for junior attorneys. The bots may soon function as quasi-employees. In the past year, more than 10 major law firms have "hired" Ross, a robotic attorney powered in part by IBM's Watson artificial intelligence, to perform legal research. Ross is designed to approximate the experience of working with a human lawyer: It can understand questions asked in normal English and provide specific, analytic answers.

Beyond helping prepare cases, AI could also predict how they'll hold up in court. Lex Machina, a company owned by LexisNexis, offers what it calls "moneyball lawyering." It applies natural-language processing to millions of court decisions to find trends that can be used to a law firm's advantage. For instance, the software can determine which judges tend to favor plaintiffs, summarize the legal strategies of opposing lawyers based on their case histories, and determine the arguments most likely to convince specific judges. A Miami-based company called

Premonition goes one step further and promises to predict the winner of a case before it even goes to court, based on statistical analyses of verdicts in similar cases. “Which attorneys win before which judges? Premonition knows,” the company says.

If you can predict the winners and losers of court cases, why not bet on them? A Silicon Valley start-up called Legalist offers “commercial litigation financing,” meaning it will pay a lawsuit’s fees and expenses if its algorithm determines that you have a good chance of winning, in exchange for a portion of any judgment in your favor. Critics fear that AI will be used to game the legal system by third-party investors hoping to make a buck.

2 | Chatbot Lawyers
Technologies like Ross and Lex Machina are intended to assist lawyers, but AI has also begun to replace them—at least in very straightforward areas of law. The most successful robolawyer yet was developed by a British teenager named Joshua Browder. Called DoNotPay, it’s a free parking-ticket-fighting chatbot that asks a series of questions about your case—Were the signs clearly marked? Were you parked illegally because of a medical emergency?—and generates a letter that can be filed with the appropriate agency. So far, the bot has helped more than 215,000 people beat traffic and parking tickets in London, New York, and Seattle. Browder recently added new functions—DoNotPay can now help people demand compensation from airlines for delayed flights and file paperwork for government housing assistance—and more are on the way.

DoNotPay is just the beginning. Until we see a major, society-changing breakthrough in artificial intelligence, robolawyers won’t dispute the finer points of copyright law or write elegant legal briefs. But chatbots could be very useful in certain types of law. Deportation, bankruptcy, and divorce disputes, for instance, typically require navigating lengthy and confusing statutes that have been interpreted in thousands of previous decisions. Chatbots could eventually analyze most every possible exception, loophole, and historical case to determine the best path forward.

As AI develops, robolawyers could help address the vast unmet legal needs of the poor. Roland Vogl, the executive director of the Stanford Program in Law, Science, and Technology, says bots will become the main entry point into the legal system. “Every legal-aid group has to turn people away because there isn’t time to process all of the cases,” he says. “We’ll see cases that get navigated through an artificially intelligent computer system, and lawyers will only get involved when it’s really necessary.” A good analogy is TurboTax: If your taxes are straightforward, you use TurboTax; if they’re not, you get an accountant. The same will happen with law.

3 | Minority Report
We’ll probably never see a court-appointed robolawyer for a criminal case, but algorithms are changing how judges mete out punishments. In many states, judges use software called COMPAS to help with setting bail and deciding whether to grant parole. The software uses information from a survey with more than 100 questions—covering things like a defendant’s gender, age, criminal history, and personal relationships—to predict whether he or she is a flight risk or likely to re-offend. The use of such software is troubling: Northpointe, the company that created COMPAS, won’t make its algorithm public, which means defense

attorneys can't bring informed challenges against judges' decisions. And a study by ProPublica found that COMPAS appears to have a strong bias against black defendants.

Forecasting crime based on questionnaires could come to seem quaint. Criminologists are intrigued by the possibility of using genetics to predict criminal behavior, though even studying the subject presents ethical dilemmas. Meanwhile, brain scans are already being used in court to determine which violent criminals are likely to re-offend. We may be headed toward a future when our bodies alone can be used against us in the criminal-justice system—even before we fully understand the biases that could be hiding in these technologies.

4 | An Explosion of Lawsuits

Eventually, we may not need lawyers, judges, or even courtrooms to settle civil disputes. Ronald Collins, a professor at the University of Washington School of Law, has outlined a system for landlord–tenant disagreements. Because in many instances the facts are uncontested—whether you paid your rent on time, whether your landlord fixed the thermostat—and the legal codes are well defined, a good number of cases can be filed, tried, and adjudicated by software. Using an app or a chatbot, each party would complete a questionnaire about the facts of the case and submit digital evidence.

"Rather than hiring a lawyer and having your case sit on a docket for five weeks, you can have an email of adjudication in five minutes," Collins told me. He believes the execution of wills, contracts, and divorces could likely be automated without significantly changing the outcome in the majority of cases.

There is a possible downside to lowering barriers to legal services, however: a future in which litigious types can dash off a few lawsuits while standing in line for a latte. Paul Ford, a programmer and writer, explores this idea of "nanolaw" in a short science-fiction story published on his website—lawsuits become a daily annoyance, popping up on your phone to be litigated with a few swipes of the finger.

Or we might see a completely automated and ever-present legal system that runs on sensors and pre-agreed-upon contracts. A company called Clause is creating "intelligent contracts" that can detect when a set of prearranged conditions are met (or broken). Though Clause deals primarily with industrial clients, other companies could soon bring the technology to consumers. For example, if you agree with your landlord to keep the temperature in your house between 68 and 72 degrees and you crank the thermostat to 74, an intelligent contract might automatically deduct a penalty from your bank account.

Experts say these contracts will increase in complexity. Perhaps one day, self-driving-car accident disputes will be resolved with checks of the vehicle's logs and programming. Your grievance against the local pizza joint's guarantee of a hot delivery in 10 minutes will be checked by a GPS sensor and a smart thermometer. Divorce papers will be prepared when your iPhone detects, through location tracking and text-message scanning, that you've been unfaithful. Your will could be executed as soon as your Fitbit detects that you're dead.

Hey, anything to avoid talking to a lawyer.

TRANSPORTATION

Our Driverless Future

Sue Halpern, *The New York Review of Books*, November 24, 2016
http://www.nybooks.com/articles/2016/11/24/driverless-intelligent-cars-road-ahead/

Editors' Note:
This book review provides a less optimistic—but more nuanced—description of autonomous cars than the previous article. It will be especially useful for the CON. It argues for example, that people will be reluctant to adopt autonomous cars, which minimizes their positive impact. In fact, semi-autonomous cars, such as Teslas have caused human drivers to become complacent behind the wheel, leading to several accidents.
~

This September, Uber, the app-summoned taxi service, launched a fleet of driverless Volvos and Fords in the city of Pittsburgh. While Google has had its own autonomous vehicles on the roads of Mountain View, California, Austin, Texas, Kirkland, Washington, and Phoenix, Arizona, for a few years, gathering data and refining its technology, Uber's Pittsburgh venture marks the first time such cars will be available to be hailed by the American public. (The world's first autonomous taxi service began offering rides in Singapore at the end of August, edging out Uber by a few weeks.)

Pittsburgh, with its hills, narrow side streets, snow, and many bridges, may not seem like the ideal venue to deploy cars that can have difficulty navigating hills, narrow streets, snow, and bridges. But the city is home to Carnegie Mellon's renowned National Robotics Engineering Center, and in the winter of 2015, Uber lured away forty of its researchers and engineers for its new Advanced Technologies Center, also in Pittsburgh, to jump-start the company's entry into the driverless car business.

Uber's autonomous vehicles have already begun picking up passengers, but they still have someone behind the wheel in the event the car hits a snag. It seems overstated to call this person a driver since much of the time the car will be driving itself. Uber's ultimate goal, and the goal of Google and Lyft and Daimler and Ford and GM and Baidu and Delphi and Mobileye and Volvo and every other company vying to bring autonomous vehicles to market, is to make that person redundant. As Hod Lipson and Melba Kurman make clear in *Driverless: Intelligent Cars and the Road Ahead*, the question is not if this can happen, but when and under what circumstances.

The timeline is a bit fuzzy. According to a remarkably bullish report issued by Morgan Stanley in 2013, sometime between 2018 and 2022 cars will have "complete autonomous capability"; by 2026, "100% autonomous penetration" of the market will be achieved. A study by the market research firm IHS Automotive predicts that by 2050, nearly all vehicles will be self-driving; a University of Michigan study says 2030. Chris Urmson, who until recently was project manager of Google's autonomous car division, is more circumspect.

"How quickly can we get this into people's hands? If you read the papers, you see maybe it's three years, maybe it's thirty years. And I am here to tell you that honestly, it's a bit of both," he told an audience at Austin's South By Southwest Festival in March. "This technology," Urmson went on, "is almost certainly going to come out incrementally. We imagine we are going to find places where the weather is good, where the roads are easy to drive—the technology might come there first. And then once we have confidence with that, we will move to more challenging locations."

As anyone looking to buy a new car these days knows, a number of technologies already cede certain tasks to the vehicle. These include windshield wipers that turn on when they sense rain, brakes that engage automatically when the car ahead is too close, blind-spot detectors, drift warnings that alert the driver when the car has strayed into another lane, cruise control that maintains a set distance from other vehicles, and the ability of the car to parallel park itself. Tesla cars go further. In "autopilot" mode they are able to steer, change lanes, and maintain proper speed, all without human intervention. YouTube is full of videos of Tesla "drivers" reading, playing games, writing, and jumping into the back seat as their cars carry on with the mundane tasks of driving. And though the company cautions drivers to keep their hands on the steering wheel when using autopilot, one of the giddiest hands-free Tesla videos was posted by Talulah Riley, the (then) wife of Elon Musk, the company's CEO.

These new technologies are sold to consumers as safety features, and it is easy to see why. Cars that slow down in relation to the vehicle in front of them don't rear-end those cars. Cars that warn drivers of lane-drift can be repositioned before they cause a collision. Cars that park themselves avoid bumping into the cars around them. Cars that sense the rain and clear it from the windshield provide better visibility. Paradoxically, though, cars equipped with these features can make driving less safe, not more, because of what Lipson and Kurman call "split responsibility." When drivers believe the car is in control, their attention often wanders or they choose to do things other than drive. When Tesla driver Joshua Brown broadsided a truck at seventy-four miles per hour last May, he was counting on the car's autopilot feature to "see" the white eighteen-wheeler turning in front of it in the bright Florida sun. It didn't, and Brown was killed. When the truck driver approached the wreckage, he told authorities, a Harry Potter video was still playing in a portable DVD player.

While Brown's collision could be considered a one-off—or, more accurately, a three-off, since there have been two other accidents involving Tesla's autopilot software—preliminary research suggests that these kinds of collisions soon may become more frequent as more cars become semiautonomous. In a study conducted by Virginia Tech in which twelve subjects were sent on a three-hour test drive around a track, 58 percent of participants in cars with lane-assist technology watched a DVD while driving and 25 percent used the time to read. Not everyone was tempted by the videos, magazines, books, and food the researchers left in the cars, but enough were that overall safety for everyone on the road diminished. As Lipson and Kurman point out, "Clearly there's a tipping point at which autonomous driving technologies will actually create more danger for human drivers rather than less." This is why Google, most prominently, is aiming to

bypass split responsibility and go directly to cars without steering wheels and brake pedals, so that humans will have no ability to drive at all.

For generations of Americans especially, and young Americans even more, driving and the open road promised a kind of freedom: the ability to light out for the territory, even if the territory was only the mall one town over. Autonomous vehicles also come with the promise of freedom, the freedom of getting places without having to pay attention to the open (or, more likely, clogged) road, and with it, the freedom to sleep, work, read e-mail, text, play, have sex, drink a beer, watch a movie, or do nothing at all. In the words of the Morgan Stanley analysts, whose enthusiasm is matched by advocates in Silicon Valley and cheerleaders in Detroit, driverless vehicles will deliver us to a "utopian society."

That utopia looks something like this: fleets of autonomous vehicles—call them taxi bots—owned by companies like Uber and Google, able to be deployed on demand, that will eliminate, for the most part, the need for private car ownership. (Currently, most privately owned cars sit idle for most of the day, simply taking up space and depreciating in value.) Fewer privately owned vehicles will result in fewer cars on the road overall. With fewer cars will come fewer traffic jams and fewer accidents. Fewer accidents will enable cars to be made from lighter materials, saving on fuel. They will be smaller, too, since cars will no longer need to be armored against one another.

With less private car ownership, individuals will be freed of car payments, fuel and maintenance costs, and insurance premiums. Riders will have more disposable income and less debt. The built environment will improve as well, as road signs are eliminated—smart cars always know where they are and where they are going—and parking spaces, having become obsolete, are converted into green spaces. And if this weren't utopian enough, the Morgan Stanley analysts estimate that switching to full vehicle autonomy will save the United States economy alone $1.3 trillion a year.

There are many assumptions embedded in this scenario, the most obvious being that people will be willing to give up private car ownership and ride in shared, driverless vehicles. (Depending on the situation, sharing either means using cars owned by fleet companies in place of privately owned vehicles, or shuttling in cars owned by fleet companies with other riders, most likely strangers going to proximate destinations.) There is no way to know yet if this will happen. In a survey by the Insurance Information Institute last May, 55 percent of respondents said they would not ride in an autonomous vehicle. But that could change as self-driving cars become more commonplace, and as today's young adults, who have been slower to get drivers' licenses and own cars than their parents' generation, and who have been early adopters of car-sharing businesses like Zipcar and Uber, become the dominant demographic.

Yet even if car-sharing membership continues its upward trajectory, these services may remain marginal. (Predictions are that 3.8 million Americans will be members of car-sharing programs by 2020—about 1 percent of the population.) And while it seemed as though the millennial generation was eschewing automobile ownership—the car industry called it the "go-nowhere" generation—its reluctance to buy cars may have had more to do with a weak economy (student

debt, unemployment, and underemployment, especially) than with desire. More millennials are buying cars now than ever before, and car sales overall are increasing. If Singapore, which has attempted to curb private car ownership by imposing heavy taxes, licensing fees, and congestion pricing, and by providing free public transit during the morning commute, is any guide, it suggests that moving the public away from private car ownership will be challenging: despite all these efforts, subway use in Singapore has gone down, not up. It is far from clear whether the country's new self-driving taxi service will be able to shift behavior and buying patterns.

There is little doubt that so far, in carefully chosen settings, where the weather is temperate, the roads are well marked, and the environment is mapped in exquisite detail, fully autonomous vehicles are safer than cars driven by humans. There were over 35,000 traffic fatalities in the United States last year, and over six million accidents, almost all due to human error. Since they were introduced in 2009, Google's self-driving cars have logged more than 1.5 million miles and have been involved in eighteen accidents, with only one considered to be the fault of the car. (At the same time, Google's human copilots have had to take over hundreds, and possibly thousands, of times.)

But Google's cars are being tested in relatively tame environments. The crucial exam begins when they are let loose, to go hither and yon, on roads without clear line markers, in snowstorms and ice storms, in heavily forested areas, and in places where GPS signals are weak or nonexistent. As Chris Urmson told that South By Southwest audience, this experiment is an unlikely prospect in the near term. But his remarks came before Baidu, Google's Chinese rival, announced it will have autonomous cars for sale in two years. It is a bold claim since so far, according to Lipson and Kurman, "to date, no robotic operating system can claim to have fully mastered three crucial capabilities: real-time reaction speed, 99.999% reliability, and better than human-level perception."

Still, the artificial intelligence guiding vehicles to full autonomy gets better and better the more miles and terrain those vehicles drive, and this is by design. As data is fed into an onboard computer from a car's many sensors and cameras, that data is parsed by an algorithm looking for statistical patterns. Those patterns enable the computer to instantaneously build a model of possible outcomes and instruct the car to proceed accordingly. The more data the algorithm has to work with—from GPS, radar, LIDAR (laser radar), sonar, inertial measurement unity (IMU), or visual clues matched to a high-definition map of topographical and geographical features—the more accurate its predictions become. The algorithm also benefits from "fleet learning": acquiring data, and thus "experience," from other vehicles equipped with the same operating system. (Autonomous cars can also communicate with each other on the road.)

This artificial intelligence enables the car to determine, for example, if the obstruction ahead is a small child or a cardboard box that has blown into the road. AI will teach it, by repetition, that the category "people" includes both those who wear trousers and those who wear skirts, those who are small and those who are tall, those who walk alone and those who walk hand-in-hand. The car will know to stop when the traffic light is red, and not to obey the green traffic light when a traffic cop is standing in the road, gesturing for it to stop.

At least that's the ideal. A car's "perception" can be stymied by mud on its cameras and by environments that do not have many distinguishing landmarks. Complex situations, like what to do when a squirrel runs into the road, followed by a dog, followed by a child, are beyond its competence. As Lipson and Kurman observe, "the technological last mile in driverless-car design is the development of software that can provide reliable artificial perception. If history is any guide, this last mile could prove to be a long haul."

There are other issues, too, that are likely to slow, though not stop, the widespread adoption of autonomous vehicles even if these technical obstacles can be overcome. Insurance is one. Now that the National Highway Traffic Safety Administration has determined that the driver of an autonomous car is its computer, will insurance need to be carried by the car manufacturer or by the software developer? (Volvo has already said it will cover the cost of accidents in its autonomous vehicles if the system has been used correctly.) Will members of car-sharing services have to waive their right to sue if a fleet car gets in an accident? And how will blame be assessed? Was the accident the fault of software that didn't accurately read the road, or the municipality that didn't maintain the road? Tort law is likely to be as challenged by the advent of self-driving cars as the automobile industry itself.

Then there are the ethical considerations. Machines can learn, but they can't process information without instructions, and as a consequence autonomous vehicles will have to be programmed in advance to respond to various life-and-death scenarios. Human drivers make such calculations all the time. They are idiosyncratic and unsystematic, two things computers are not, which is why robotic cars are safer than cars driven by humans.

But there are ethical precepts embedded in artificial intelligence. Cars will be programmed to execute a predetermined calculus that puts a value on the living creatures that come into their path. Will it be utilitarian, and steer the car in the direction that will kill or injure the least number of pedestrians, or will there be other rules—when possible, spare children, for instance? Someone will have to write the code to set these limits, and it is not clear yet who will tell the coders what to write. Will we have a referendum, or will these decisions be made by engineers or corporate executives? If we needed a reminder that autonomous vehicles are actually robots, this may be it.

At the moment, it is easy not to notice. Cars with semiautonomous features look like other cars on the road. Most likely, no one passing the Audi Q5 that drove itself (for the most part) from San Francisco to New York last March knew it was being piloted by GPS and an array of sensors, cameras, and algorithms. This will change when steering wheels and brake pedals are museum pieces, when cars are made of carbon fiber that has been punched out by 3D printers, and when they display the names of tech companies like video card maker NVIDIA and optical sensor pioneer Mobileye, rather than GM and Chrysler, on their grilles—if they have grilles.

The major car makers, rushing to make alliances with tech companies, understand their days of dominance are numbered. "We are rapidly becoming both an auto company and a mobility company," Bill Ford, the chairman of Ford Motor Company, told an audience in Kansas City in February. He knows that if the fleet model prevails, Ford and other car manufacturers will be

selling many fewer cars. More crucially, the winners in this new system will be the ones with the best software, and the best software will come from the most robust data, and the companies with the most robust data are the tech companies that have been hoovering it up for years: Google most of all.

"The mobility revolution is going to affect all of us personally and many of us professionally," Ford said that day in Kansas City. He might have been thinking about car salespeople, whose jobs are likely to become obsolete, but before that it will be the taxi drivers and truckers who will be displaced by vehicles that drive themselves. Historically these have been the jobs that have provided incomes to recently arrived immigrants and to people without college degrees. Without them yet another trajectory into the middle class will be eliminated.

What of Uber drivers themselves? These are the poster people for the gig-economy, "entrepreneurs"—which is to say freelancers—who use their own cars to ferry people around. "Obviously the self-driving car thing is freaking people out a little bit," an Uber driver in Pittsburgh named Ryan told a website called TechRepublic. And, he went on, he learned about Uber's plans from the media, not from the company. "If it's a negative thing, they let you find out for yourself." As media critic Douglas Rushkoff has written, "Uber's drivers are the R&D for Uber's driverless future. They are spending their labor and capital investments (cars) on their own future unemployment."

All economies have winners and losers. It does not take a sophisticated algorithm to figure out that the winners in the decades ahead are going to be those who own the robots, for they will have vanquished labor with their capital. In the case of autonomous vehicles, a few companies are now poised to control a necessary public good, the transportation of people to and from work, school, shopping, recreation, and other vital activities. This salient fact is often lost in the almost unanimously positive reception of the coming "mobility revolution," as Bill Ford calls it.
Also lost is this: the most optimistic estimates of the safety and environmental benefits of the transition to fleet-owned autonomous vehicles, and the ones often used to tout it, are based on models derived from cities with high-capacity public transit systems. Obviously there are many cities without such systems. Where they exist, according to the Boston Consulting Group, it would take only two passengers sharing an autonomous taxi to make the per-passenger cost comparable to that of mass transit. Perhaps not surprisingly, lawmakers in this country are now using the autonomous vehicle future laid out by companies like Uber and Google to block investment in mass transit.

But why take a train or a bus to a central location on a fixed schedule when you can take a car, at will, to the exact place you want to go? One can imagine Google offering rides for free as long as passengers are willing to "share" the details of where they are going, what they are buying, who they are with, and which products their eyes are drawn to on the ubiquitous (but targeted!) ads that are playing in the car's cabin. Like websites that won't load if you block ads or disallow cookies, or like Gmail, which does not allow users to opt out of having their mail read by the company's automated scanners, one can also imagine feeling as if one had no choice: give up your data or take a hike.

Lipson and Kurman worry that the software driving driverless cars, like all software, will be ripe for hacking and sabotage. While this is not a concern to be taken lightly, especially after a pair of hacker/researchers were able to take remote control of a Jeep Cherokee speeding down the highway last year, this is an engineering problem, with an engineering solution. (To be clear: the hack was an experiment.) More worrisome is the authors' suggestion that autonomous vehicles, with their high-definition cameras and sensors, could "morph into ubiquitous robotic spies," sending information to intelligence agencies and law enforcement departments, among others, about people inside and outside the car, tracking anyone, anywhere. Welcome to utopia, where cars are autonomous, but their passengers not so much.

HEALTH CARE

How AI Will Cure America's Sick Health Care System

Kevin Maney, *Newsweek*, May 24, 2017
http://www.newsweek.com/2017/06/02/ai-cure-america-sick-health-care-system-614583.html

For decades, technology has relentlessly made phones, laptops, apps and entire industries cheaper and better—while health care has stubbornly loitered in an alternate universe where tech makes everything more expensive and more complex.

Now startups are applying artificial intelligence (AI), floods of data and automation in ways that promise to dramatically drive down the costs of health care while increasing effectiveness. If this profound trend plays out, within five to 10 years, Congress won't have to fight about the exploding costs of Medicaid and insurance. Instead, it might battle over what to do with a massive windfall. Today's debate over the repeal of Obamacare would come to seem as backward as a discussion about the merits of leeching.

Hard to believe? One proof point is in the maelstrom of activity around diabetes, the most expensive disease in the world. In the U.S., nearly 10 percent of the population has diabetes, around 30 million people. Within a decade, some experts say, the number of diabetics in China will outnumber the entire U.S. population. Most people who suffer from the disease spend $5,000 to $10,000 a year on medication, and diabetics with complications can spend hundreds of thousands of dollars on doctor and hospital bills. That and the lost wages of diabetics cost the U.S. alone more than $245 billion a year, according to the Centers for Disease Control and Prevention.

That's an enormous problem to solve — and a pile of potential cash and customers to be won— which is why diabetes is attracting entrepreneurs like ants to a dropped ice cream cone. One of those entrepreneurs is Sami Inkinen. He was a co-founder of the real estate site Trulia and has long been an endurance athlete, competing seriously in triathlons and Ironman events. In 2014, he and his wife rowed from California to Hawaii. None of this fits the typical profile of a diabetic, yet in 2011, soon after yet another triathlon, Inkinen was diagnosed with Type 2 diabetes. And like many driven, super-smart data geeks, he dove into research to understand everything about his condition.

That journey led him to Dr. Stephen Phinney, a medical researcher at the University of California, Davis, and Jeff Volek, a scientist at Ohio State. Phinney and Volek wrote two books together about low-carbohydrate diets and published scientific papers describing how constant adjustments to diet and lifestyle can reverse diabetes in many patients. Diabetes is almost never treated that way because the program is too hard for most people to stick to. It requires so much coaching and scrutiny by medical professionals, you'd pretty much have to hire a live-in doctor. Inkinen convinced Phinney and Volek that technology could essentially re-create a live-in doctor and diabetes coach in a smartphone. Together, the three founded Virta Health in 2014. The company stayed in stealth mode until now, launching in March. "It felt like a duty to do this,"

Inkinen tells Newsweek . “Here is an epidemic of epic proportion, and nothing is working. We can combine science and technology to solve the problem at much lower cost and do it safely.” Here’s how Virta works and why its approach is so important to the future of health care. On the front end, Virta is software on a smartphone. Diabetics who sign up agree to regularly enter data: glucose levels, weight, blood pressure, activity. Some do this by manually entering information; others use devices like a Fitbit or connected scales to automatically send it in. The app also frequently asks multiple-choice questions about mood, energy levels and hunger — more data that the AI software crunches to learn about the patient, look for warning signs and symptoms and guide Virta’s doctors.

On the back end, Virta hires doctors who get streams of updates from Virta’s software and use the data to help them make decisions about how to adjust each patient’s diet and medications or anything else that might affect that person’s health. “Any clinical decision is always made by a doctor,” Inkinen says. “But the software increases productivity by 10-X.” (That’s 10 times, in Silicon Valley–speak.) When all this works and the patient follows the program’s strict dietary and medical controls, diabetes can be reversed, clinical trials of Virta’s system have shown. Around 87 percent of patients who had been relying on insulin to control their condition either decreased their dose or eliminated their use of insulin completely—a success rate that matches that for bariatric surgery, which is an expensive, invasive, last-ditch effort for severe diabetics.

Virta leverages AI software, smartphones and cloud computing to allow its doctors to continually interact with many times more patients than they can in a clinic or hospital, and it gives its diabetic patients a cross between a pocket doctor and a guardian angel. The result is a promising treatment for diabetes that could get many sufferers off medication and keep them out of doctors’ offices and hospital emergency rooms. And that, in turn, would greatly lower the overall cost of diabetes.

Virta is just one startup of many attacking diabetes. Livongo is a more automated but less doctor-oriented version of Virta’s program. The company, which raised $52.5 million in March, makes a wireless glucose-reading device that uploads the diabetic’s data. AI software learns about the patient and sends a stream of tips and information intended to help the diabetic manage the disease and stay out of hospitals. Yet another new startup, Fractyl, takes a more medical approach. It invented a type of catheter that seems to cause changes in the intestines that result in reversing diabetes.

Startup tracker Crunchbase lists about 130 new tech-oriented companies (the number changes constantly) involved in some aspect of diabetes. While many of these startups will fail, it’s hard to imagine that some won’t have a significant impact.

These efforts matter to all of us because diabetes is such an enormous drain on health care resources. Venture capitalist Hemant Taneja, who helped start Livongo, says technology could take $100 billion out of the annual cost of diabetes in the U.S. Imagine if even 20 percent of diabetics could get off medication and have little need for a doctor’s care. All of those medical resources would get freed up for other patients and other conditions, which should help lower prices of health care for all. “If we want to massively lower health care costs, we need to figure

out how to address metabolic health issues [like diabetes] at their core," Inkinen says. "I would bet my house that in 15 years, the future health care company looks like what we're doing today—not treating diseases at the end of the road but catching them along the way and reversing them."

'Alexa, What's Wrong With Me?'

Over the past decade, medical records in the U.S.—long kept on paper in doctors' horrible handwriting—have been digitized and fed into software. That hasn't helped lower health care costs yet, and in fact it is adding to them as systems get installed and medical professionals learn to use software that can be clunky. Epic Systems, the biggest electronic medical records company, handles 54 percent of patients' records in the U.S. but gets bad marks for being so hard to use that it eats up doctors' and nurses' time. One report from Becker's Hospital Review said that almost 30 percent of Epic clients wouldn't recommend it to their peers. A survey by Black Book Market Research found that 30 percent of hospital personnel were dissatisfied with their EMR systems, with Epic getting the strongest dissatisfaction.

But there's a larger gain from the pain of EMRs: Enormous amounts of medical information are now digitized. As more medical interaction happens online—as with Virta or Livongo—the more kinds of data we'll collect. Internet of Things devices, whether Fitbits or connected glucose meters or potential new devices like Apple AirPods that take biometric readings, will add yet more data. All this data can help AI software learn about diseases in general, and about individual patients, opening up new ways for technology to be applied.

Some of the new applications of AI will simply improve a tragically inefficient health care industry. Qventus is a startup using AI to take all the data flowing through a hospital to learn how to free up doctors and nurses to see more patients and improve outcomes. "We're creating efficiency out of seemingly nothing," Qventus CEO Mudit Garg tells me. "Two years ago, work like this was so unsexy. But this is where the rubber meets the road."

One of his clients, Mercy Hospital Fort Smith in Fort Smith, Arkansas, has been able to treat 3,000 more patients a year with the same resources, an increase of 18 percent. Here again, technology is increasing the supply of medical services, potentially changing the cost equation that keeps forcing health care prices higher.

AI is also starting to automate some of the work of doctors. IBM's Watson, which uses machine learning and massive computing power to reason its way through questions, is on its way to becoming the best diagnostician on the planet. Its software can soak up all manner of available (and anonymized) patient data, plus the tens of thousands of medical research papers published every year (far more than any human could read). The system can even keep up with the news, learning, for instance, which regions are affected by a certain contagious disease, which might help diagnose someone who recently traveled to one of those areas. By asking patients a series of questions spoken into any kind of computer or connected device, Watson can quickly narrow down the possible causes of a medical problem. Today, IBM works on test projects with major hospitals like the Cleveland Clinic to put Watson in the hands of doctors, who are learning how to use the technology like a brilliant assistant.

But the day will come when Watson or something like it is available to everyone through a smartphone or some other device. Amazon is starting down that path by partnering with HealthTap to offer what it calls Dr. A.I. on Alexa, Amazon's voice-activated AI gadget for consumers. It's not nearly as robust as Watson but works on the same idea. Just tell it your medical problem, and it will ask you questions to help narrow down what it might be.

As health care AI develops, startups are also creating new kinds of genomics-based medicine. Just 16 years ago, the Human Genome Project and geneticist Craig Venter's startup, Celera Genomics, published the results of their human genome sequencing within a day of each other in 2001. Venter said his project took 20,000 hours of processor time on a supercomputer. This year, startup Color Genomics is offering a $249 genetic test that can sequence most of the pertinent genes in the human body. Color's goal is to make genetic sequencing so cheap and easy that every baby born will have it done, and the data will inform his or her health care for life. Combine genetic data about a person with all the kinds of data Watson can ingest, and we're close to being able to build AI software that can at least supplant that first visit to a doctor when you're sick—which, of course, is when you least want to travel to a doctor's office. Instead, people will increasingly speak to a smartphone or to something like Dr. A.I. on Alexa about their health problems and, if necessary, send in photos of that rash or funky toe. If the system has your health care records and genetic data, it can gain more insight into your condition than any doctor operating on an informed hunch.

On many occasions, the app might tell the user the problem is nothing serious—a robot equivalent of "Take two aspirin and call me in the morning." Other times, the app might send the user to a clinic to get a test or X-ray. If that's how it plays out, a large chunk of the traffic into doctors' offices and hospitals will fade away.

Add it up, and in these next few years we're going to see a parade of tech applications that reduce demand on the health care system while giving all of us more access to care. Doctors should be freed up to do a better job for patients who truly need their attention. Theoretically, all of this will help keep more people healthier. And if we're all healthier and using health care less, the laws of supply and demand should kick in, sending the overall cost of health care tumbling. However, there are bumps ahead because, as our erudite president recently said, "nobody knew that health care could be so complicated."

The Automation Rx

The economics of health care are weird. First of all, the usual forces don't apply to highly regulated industries, and health care is perhaps the most regulated in the U.S. and around the world because lives are at stake. In most countries, regulators prevent AI software from crossing the line into independently offering a diagnosis or clinical advice—that's strictly the purview of doctors. New medical devices, like Fractyl's, have to get approval from the Food and Drug Administration. Lobbyists often slow regulatory change to maintain the status quo and benefit incumbents charging inflated prices.

Personal health care decisions in the U.S. often get influenced by insurance companies, employers who pay for health benefits, and Medicare. Unlike most industries, consumers in health care don't have much information about pricing or quality, so they can't weigh options and make rational choices. Moreover, we think about health differently from anything else we buy. Many of us are never satiated with health care—we always want more and better health care, if we can afford it. One study published in March showed that telehealth —making doctors available by video call—prompted people to seek care for minor illnesses they otherwise would've ignored. Only 12 percent of telehealth "visits" replaced in-person visits, and the other 88 percent was new demand.

Until recently, most new medical technology has been high-end products that give doctors and hospitals a reason to charge more for something that couldn't have been done in the past. Think MRI machines or robotic limbs. These improve quality of life but add to costs. In 2008, the Congressional Budget Office concluded, "The most important factor driving the long-term growth of health care costs has been the emergence, adoption, and widespread diffusion of new medical technologies and services."

The next wave of health care technology is different. The combination of data and AI was not available until the past year or two, and it can lead to the kind of automation that has disrupted so many other industries. Many health care entrepreneurs are focused precisely on the win-win-win prospect of lowering the cost of care while making it better and available to more people. Of course, there will be challenges to address, such as making sure our highly sensitive medical data stays protected and private, even as it flies around various networks and systems.

As startups bring these technologies online, they're often doing an end run around insurance companies, instead finding demand among consumers or employers who offer health coverage. Livongo, for instance, points out to companies that each diabetic employee costs thousands of dollars a year in care. Pay for the Livongo service, the pitch goes, and your company will save money as those employees better manage their conditions. By last year, Livongo had signed up more than 50 large customers, including Quicken, Office Depot, Office Max and S.C. Johnson & Son. As the thinking goes among health care startups, once employers and consumers embrace new technology, insurance companies, regulators and health care incumbents will have to follow. As that happens, the technologists promise, economic forces will finally stall or reverse the climbing cost of health care in the U.S. and around the world, a development that would, if we're lucky, leave the president and just about every member of Congress speechless.

MILITARY

Elon Musk Leads 116 Experts Calling for Outright Ban of Killer Robots

Samuel Gibbs, *The Guardian*, August 20, 2017.
https://www.theguardian.com/technology/2017/aug/20/elon-musk-killer-robots-experts-outright-ban-lethal-autonomous-weapons-war

Editors' Note:
The development of weapons systems that incorporate AI raises several ethical and military-operational questions. This article looks at possible bans on the production or deployment of AI capable weapons.

It is also important to note that very few, if any, fully-autonomous weapons are currently deployed by any military in the world. Many weapons systems in use are partly-autonomous, but they are still controlled by humans. For example, U.S. Predator drones take-off and land using AI software, but they are remotely piloted by humans, and humans make the decision when to fire.

For more impacts related military issues, see the section on existential risks.
~

Some of the world's leading robotics and artificial intelligence pioneers are calling on the United Nations to ban the development and use of killer robots.

Tesla's Elon Musk and Alphabet's Mustafa Suleyman are leading a group of 116 specialists from across 26 countries who are calling for the ban on autonomous weapons.

The UN recently voted to begin formal discussions on such weapons which include drones, tanks and automated machine guns. Ahead of this, the group of founders of AI and robotics companies have sent an open letter to the UN calling for it to prevent the arms race that is currently under way for killer robots.

In their letter, the founders warn the review conference of the convention on conventional weapons that this arms race threatens to usher in the "third revolution in warfare" after gunpowder and nuclear arms.

The founders wrote: "Once developed, lethal autonomous weapons will permit armed conflict to be fought at a scale greater than ever, and at timescales faster than humans can comprehend. These can be weapons of terror, weapons that despots and terrorists use against innocent populations, and weapons hacked to behave in undesirable ways.

"We do not have long to act. Once this Pandora's box is opened, it will be hard to close."

Experts have previously warned that AI technology has reached a point where the deployment of autonomous weapons is feasible within years, rather than decades. While AI can be used to make the battlefield a safer place for military personnel, experts fear that offensive weapons that operate on their own would lower the threshold of going to battle and result in greater loss of human life.

The letter, launching at the opening of the International Joint Conference on Artificial Intelligence (IJCAI) in Melbourne on Monday, has the backing of high-profile figures in the robotics field and strongly stresses the need for urgent action, after the UN was forced to delay a meeting that was due to start Monday to review the issue.

The founders call for "morally wrong" lethal autonomous weapons systems to be added to the list of weapons banned under the UN's convention on certain conventional weapons (CCW) brought into force in 1983, which includes chemical and intentionally blinding laser weapons.

Toby Walsh, Scientia professor of artificial intelligence at the University of New South Wales in Sydney, said: "Nearly every technology can be used for good and bad, and artificial intelligence is no different. It can help tackle many of the pressing problems facing society today: inequality and poverty, the challenges posed by climate change and the ongoing global financial crisis.

"However, the same technology can also be used in autonomous weapons to industrialise war. We need to make decisions today choosing which of these futures we want."

Musk, one of the signatories of the open letter, has repeatedly warned for the need for pro-active regulation of AI, calling it humanity's biggest existential threat, but while AI's destructive potential is considered by some to be vast it is also thought be distant.

Ryan Gariepy, the founder of Clearpath Robotics said: "Unlike other potential manifestations of AI which still remain in the realm of science fiction, autonomous weapons systems are on the cusp of development right now and have a very real potential to cause significant harm to innocent people along with global instability."

This is not the first time the IJCAI, one of the world's leading AI conferences, has been used as a platform to discuss lethal autonomous weapons systems. Two years ago the conference was used to launch an open letter signed by thousands of AI and robotics researchers including Musk and Stephen Hawking similarly calling for a ban, which helped push the UN into formal talks on the technologies.

The UK government opposed such a ban on lethal autonomous weapons in 2015, with the Foreign Office stating that "international humanitarian law already provides sufficient regulation for this area". It said that the UK was not developing lethal autonomous weapons and that all weapons employed by UK armed forces would be "under human oversight and control".

Science fiction or science fact?

While the suggestion of killer robots conjures images from science fiction such as the Terminator's T-800 or Robocop's ED-209, lethal autonomous weapons are already in use.

Samsung's SGR-A1 sentry gun, which is reportedly technically capable of firing autonomously but is disputed whether it is deployed as such, is in use along the South Korean border of the 2.5m-wide Korean Demilitarized Zone.

The fixed-place sentry gun, developed on behalf of the South Korean government, was the first of its kind with an autonomous system capable of performing surveillance, voice-recognition, tracking and firing with mounted machine gun or grenade launcher. But it is not the only autonomous weapon system in development, with prototypes available for land, air and sea combat.

The UK's Taranis drone, in development by BAE Systems, is intended to be capable of carrying air-to-air and air-to-ground ordnance intercontinentally and incorporating full autonomy. The unmanned combat aerial vehicle, about the size of a BAE Hawk, the plane used by the Red Arrows, had its first test flight in 2013 and is expected to be operational some time after 2030 as part of the Royal Air Force's Future Offensive Air System, destined to replace the human-piloted Tornado GR4 warplanes.

Russia, the US and other countries are currently developing robotic tanks that can either be remote controlled or operate autonomously. These projects range from autonomous versions of the Russian Uran-9 unmanned combat ground vehicle, to conventional tanks retrofitted with autonomous systems.

The US's autonomous warship, the Sea Hunter built by Vigor Industrial, was launched in 2016 and, while still in development, is intended to have offensive capabilities including anti-submarine ordnance. Under the surface, Boeing's autonomous submarine systems built on the Echo Voyager platform are also being considered for long-range deep-sea military use.

ISSUES

EMPLOYMENT, INEQUALITY, AND THE FUTURE OF WORK

How can we address real concerns over artificial intelligence?

Harry Armstrong and Jared Robert Keller, *The Guardian,* Sept 15, 2016.
https://www.theguardian.com/media-network/2016/sep/15/responsiblity-real-concerns-artificial-intelligence-technology

Editors' Note:
One of the things to take away from this article is that it compares fears and concerns that people have about AI to similar concerns that arose in during the mid 20th century about automation. In doing so it explains why there are important concerns about AI, but that discourse around AI needs to remain realistic, rather than alarmist about these possible harms. The key, according to this article, is to ensure that AI is developed responsibly, rather than opposing the development of AI technology altogether.
~

The cashiers' demands were simple: management must remove the tracking software they had installed in the checkout terminals, or the cashiers would refuse to return to work. The technology that tracked their every movement – their speed, efficiency, etc – had been installed without their knowledge, they asserted, and was an invasion of privacy.

While this seems like something one might see today, this happened nearly 40 years ago at the biggest supermarket chain in Denmark. It is easy to see these cashiers as luddites, anti-progress and anti-technology, but like the original luddites they had valid concerns about the way that new technologies are used by employers. New technologies can not only lead to jobs lost through automation, but can also change the nature of work itself in detrimental ways.

Walter Reuther, president of the United Automobile Workers union from 1946-1970, raised similar concerns about the impact of automation and decision-making machines on jobs when he appeared before the US Congress in 1955 – 20 years before the Danish cashier walkout. "We believe that we have got to look at this problem realistically, with honesty, and with courage", Reuther told the Congressional committee: "When you say there is a problem here it doesn't mean that you are opposed to automation. It merely means that you are trying to anticipate the problem so that we can meet it in advance."

As we confront similar questions around artificial intelligence (AI), we must distinguish between fear of new technologies and concern about their implications. The latter plays an important role in cultivating the right conversations to ensure that new technologies are deployed ethically and responsibly.

Emerging technologies like AI present us with many opportunities to improve the way we work, to provide better services and products in more efficient ways, and to do things we have never been able to do in the past. If we do not acknowledge and take on board people's valid concerns, we risk seeing the potential benefits of these technologies lost under a mountain of fear and negative press. As a result, we could lose public trust.

It is easy to see how this could happen. In the US, criminal "risk assessments" based on predictive analytics have already been shown to be biased against black people because the data used to build the system was inherently biased, and a more recent evaluation of Chicago's use of predictive policing has shown that the system doesn't help reduce homicides (as it was designed to do).

We need to have a mature, informed and inclusive conversation about the future of automation and the potential impact of new technologies. […]

Machines have been taking over tasks from human workers for centuries, and for nearly as long, people have been discussing, debating, and arguing over how to respond. Yet these fears keep recurring. Even those generations that recognise the recurrent nature of these fears cannot resist making similar doomsday predictions about mass unemployment or the end of work.

One reason for this seems to be that each new generation believes it is confronting fundamental technological advances that far surpass those dealt with by any other generation. Each new generation manages to convince itself that they are the ones living in the age when the exception finally proves the rule. The AI story is often painted with the same brush, particularly as it increasingly able to take over cognitive and decision-making tasks. Is AI the exception to the rule? It depends who you talk to but even if AI isn't truly a fundamental advance in this sense, it doesn't mean that it won't have a profound impact on society.

In fact it already is. As AI finds its way into all sorts of areas, automating or supplementing existing jobs, it is having a fundamental impact on the way we live – from online advertising to credit scores or the collaboration between Google DeepMind and the NHS.

So, it is now that we need to have a conversation about the immediate and future implications of these new technologies, before concern and fear take over. Looking forward and making informed predictions about things like which skills might be important in the future will help us prepare the next generation, as best we can, for any disruptions ahead.

From a public policy point of view, while we don't have any proper regulation around the use of AI, some steps are being taken: the UK government and HP are developing ethical frameworks for the use of data and AI. This is a good way to start to develop regulation and public trust but to take this further it is crucial we also have the right institutions.

This is where something like a dedicated machine intelligence commission could come into play. A new public institution like this would support an informed public dialogue, help the responsible development of new generations of algorithms, machine learning tools and uses of big data, and ensure that the public interest is protected for future generations.

[…]

Automation and anxiety

The Economist, June 25, 2016.
https://www.economist.com/news/special-report/21700758-will-smarter-machines-cause-mass-unemployment-automation-and-anxiety

Editors' Note:
This article answers looks at the question of whether or not AI will cause mass unemployment and argues that it will not. Most of the article explains why economic changes do not necessarily lead to mass unemployment and it includes many examples from the past to back up this claim. The article also explains how new jobs can be created due to the economic benefits of implementing AI.

For greater analysis on government policy related to the challenges discussed in this article see the section on macroeconomics.
~

Sitting in an office in San Francisco, Igor Barani calls up some medical scans on his screen. He is the chief executive of Enlitic, one of a host of startups applying deep learning to medicine, starting with the analysis of images such as X-rays and CT scans. It is an obvious use of the technology. Deep learning is renowned for its superhuman prowess at certain forms of image recognition; there are large sets of labelled training data to crunch; and there is tremendous potential to make health care more accurate and efficient.

Dr Barani (who used to be an oncologist) points to some CT scans of a patient's lungs, taken from three different angles. Red blobs flicker on the screen as Enlitic's deep-learning system examines and compares them to see if they are blood vessels, harmless imaging artefacts or malignant lung nodules. The system ends up highlighting a particular feature for further investigation. In a test against three expert human radiologists working together, Enlitic's system was 50% better at classifying malignant tumours and had a false-negative rate (where a cancer is missed) of zero, compared with 7% for the humans. Another of Enlitic's systems, which examines X-rays to detect wrist fractures, also handily outperformed human experts. The firm's technology is currently being tested in 40 clinics across Australia.

A computer that dispenses expert radiology advice is just one example of how jobs currently done by highly trained white-collar workers can be automated, thanks to the advance of deep learning and other forms of artificial intelligence. The idea that manual work can be carried out by machines is already familiar; now ever-smarter machines can perform tasks done by information workers, too. What determines vulnerability to automation, experts say, is not so much whether the work concerned is manual or white-collar but whether or not it is routine. Machines can already do many forms of routine manual labour, and are now able to perform some routine cognitive tasks too. As a result, says Andrew Ng, a highly trained and specialised radiologist may now be in greater danger of being replaced by a machine than his own executive

assistant: “She does so many different things that I don’t see a machine being able to automate everything she does any time soon.”

So which jobs are most vulnerable? In a widely noted study published in 2013, Carl Benedikt Frey and Michael Osborne examined the probability of computerisation for 702 occupations and found that 47% of workers in America had jobs at high risk of potential automation. In particular, they warned that most workers in transport and logistics (such as taxi and delivery drivers) and office support (such as receptionists and security guards) “are likely to be substituted by computer capital”, and that many workers in sales and services (such as cashiers, counter and rental clerks, telemarketers and accountants) also faced a high risk of computerisation. They concluded that “recent developments in machine learning will put a substantial share of employment, across a wide range of occupations, at risk in the near future.” Subsequent studies put the equivalent figure at 35% of the workforce for Britain (where more people work in creative fields less susceptible to automation) and 49% for Japan. […]

And this is only the start. “We are just seeing the tip of the iceberg. No office job is safe,” says Sebastian Thrun, an AI professor at Stanford known for his work on self-driving cars. Automation is now “blind to the colour of your collar”, declares Jerry Kaplan, another Stanford academic and author of “Humans Need Not Apply”, a book that predicts upheaval in the labour market. Gloomiest of all is Martin Ford, a software entrepreneur and the bestselling author of “Rise of the Robots”. He warns of the threat of a “jobless future”, pointing out that most jobs can be broken down into a series of routine tasks, more and more of which can be done by machines.

In previous waves of automation, workers had the option of moving from routine jobs in one industry to routine jobs in another; but now the same “big data” techniques that allow companies to improve their marketing and customer-service operations also give them the raw material to train machine-learning systems to perform the jobs of more and more people. “E-discovery” software can search mountains of legal documents much more quickly than human clerks or paralegals can. Some forms of journalism, such as writing market reports and sports summaries, are also being automated.

Predictions that automation will make humans redundant have been made before, however, going back to the Industrial Revolution, when textile workers, most famously the Luddites, protested that machines and steam engines would destroy their livelihoods. […]

As computers began to appear in offices and robots on factory floors, President John F. Kennedy declared that the major domestic challenge of the 1960s was to “maintain full employment at a time when automation…is replacing men”. In 1964 a group of Nobel prizewinners, known as the Ad Hoc Committee on the Triple Revolution, sent President Lyndon Johnson a memo alerting him to the danger of a revolution triggered by “the combination of the computer and the automated self-regulating machine”. This, they said, was leading to a new era of production “which requires progressively less human labour” and threatened to divide society into a skilled

elite and an unskilled underclass. The advent of personal computers in the 1980s provoked further hand-wringing over potential job losses.

Yet in the past technology has always ended up creating more jobs than it destroys. That is because of the way automation works in practice, explains David Autor, an economist at the Massachusetts Institute of Technology. Automating a particular task, so that it can be done more quickly or cheaply, increases the demand for human workers to do the other tasks around it that have not been automated.

There are many historical examples of this in weaving, says James Bessen, an economist at the Boston University School of Law. During the Industrial Revolution more and more tasks in the weaving process were automated, prompting workers to focus on the things machines could not do, such as operating a machine, and then tending multiple machines to keep them running smoothly. This caused output to grow explosively. In America during the 19th century the amount of coarse cloth a single weaver could produce in an hour increased by a factor of 50, and the amount of labour required per yard of cloth fell by 98%. This made cloth cheaper and increased demand for it, which in turn created more jobs for weavers: their numbers quadrupled between 1830 and 1900. In other words, technology gradually changed the nature of the weaver's job, and the skills required to do it, rather than replacing it altogether.

In a more recent example, automated teller machines (ATMs) might have been expected to spell doom for bank tellers by taking over some of their routine tasks, and indeed in America their average number fell from 20 per branch in 1988 to 13 in 2004, Mr Bessen notes. But that reduced the cost of running a bank branch, allowing banks to open more branches in response to customer demand. The number of urban bank branches rose by 43% over the same period, so the total number of employees increased. Rather than destroying jobs, ATMs changed bank employees' work mix, away from routine tasks and towards things like sales and customer service that machines could not do.

The same pattern can be seen in industry after industry after the introduction of computers, says Mr Bessen: rather than destroying jobs, automation redefines them, and in ways that reduce costs and boost demand. In a recent analysis of the American workforce between 1982 and 2012, he found that employment grew significantly faster in occupations (for example, graphic design) that made more use of computers, as automation sped up one aspect of a job, enabling workers to do the other parts better. The net effect was that more computer-intensive jobs within an industry displaced less computer-intensive ones. Computers thus reallocate rather than displace jobs, requiring workers to learn new skills. This is true of a wide range of occupations, Mr Bessen found, not just in computer-related fields such as software development but also in administrative work, health care and many other areas. Only manufacturing jobs expanded more slowly than the workforce did over the period of study, but that had more to do with business cycles and offshoring to China than with technology, he says.

So far, the same seems to be true of fields where AI is being deployed. For example, the introduction of software capable of analysing large volumes of legal documents might have been expected to reduce the number of legal clerks and paralegals, who act as human search engines during the "discovery" phase of a case; in fact automation has reduced the cost of discovery and increased demand for it. "Judges are more willing to allow discovery now, because it's cheaper and easier," says Mr Bessen. The number of legal clerks in America increased by 1.1% a year between 2000 and 2013. Similarly, the automation of shopping through e-commerce, along with more accurate recommendations, encourages people to buy more and has increased overall employment in retailing. In radiology, says Dr Barani, Enlitic's technology empowers practitioners, making average ones into experts. Rather than putting them out of work, the technology increases capacity, which may help in the developing world, where there is a shortage of specialists.

And while it is easy to see fields in which automation might do away with the need for human labour, it is less obvious where technology might create new jobs. "We can't predict what jobs will be created in the future, but it's always been like that," says Joel Mokyr, an economic historian at Northwestern University. Imagine trying to tell someone a century ago that her great-grandchildren would be video-game designers or cybersecurity specialists, he suggests. "These are jobs that nobody in the past would have predicted."

Similarly, just as people worry about the potential impact of self-driving vehicles today, a century ago there was much concern about the impact of the switch from horses to cars, notes Mr Autor. Horse-related jobs declined, but entirely new jobs were created in the motel and fast-food industries that arose to serve motorists and truck drivers. As those industries decline, new ones will emerge. Self-driving vehicles will give people more time to consume goods and services, increasing demand elsewhere in the economy; and autonomous vehicles might greatly expand demand for products (such as food) delivered locally.

There will also be some new jobs created in the field of AI itself. Self-driving vehicles may need remote operators to cope with emergencies, or ride-along concierges who knock on doors and manhandle packages. Corporate chatbot and customer-service AIs will need to be built and trained and have dialogue written for them (AI firms are said to be busy hiring poets); they will have to be constantly updated and maintained, just as websites are today. And no matter how advanced artificial intelligence becomes, some jobs are always likely to be better done by humans, notably those involving empathy or social interaction. Doctors, therapists, hairdressers and personal trainers fall into that category. An analysis of the British workforce by Deloitte, a consultancy, highlighted a profound shift over the past two decades towards "caring" jobs: the number of nursing assistants increased by 909%, teaching assistants by 580% and careworkers by 168%.

Focusing only on what is lost misses "a central economic mechanism by which automation affects the demand for labour", notes Mr Autor: that it raises the value of the tasks that can be done only by humans. Ultimately, he says, those worried that automation will cause mass

unemployment are succumbing to what economists call the "lump of labour" fallacy. "This notion that there's only a finite amount of work to do, and therefore that if you automate some of it there's less for people to do, is just totally wrong," he says. Those sounding warnings about technological unemployment "basically ignore the issue of the economic response to automation", says Mr Bessen.

But couldn't this time be different? As Mr Ford points out in "Rise of the Robots", the impact of automation this time around is broader-based: not every industry was affected two centuries ago, but every industry uses computers today. During previous waves of automation, he argues, workers could switch from one kind of routine work to another; but this time many workers will have to switch from routine, unskilled jobs to non-routine, skilled jobs to stay ahead of automation. That makes it more important than ever to help workers acquire new skills quickly. But so far, says Mr Autor, there is "zero evidence" that AI is having a new and significantly different impact on employment. And while everyone worries about AI, says Mr Mokyr, far more labour is being replaced by cheap workers overseas.

Another difference is that whereas the shift from agriculture to industry typically took decades, software can be deployed much more rapidly. Google can invent something like Smart Reply and have millions of people using it just a few months later. Even so, most firms tend to implement new technology more slowly, not least for non-technological reasons. Enlitic and other companies developing AI for use in medicine, for example, must grapple with complex regulations and a fragmented marketplace, particularly in America (which is why many startups are testing their technology elsewhere). It takes time for processes to change, standards to emerge and people to learn new skills. "The distinction between invention and implementation is critical, and too often ignored," observes Mr Bessen.

What of the worry that new, high-tech industries are less labour-intensive than earlier ones? Mr Frey cites a paper he co-wrote last year showing that only 0.5% of American workers are employed in industries that have emerged since 2000. "Technology might create fewer and fewer jobs, while exposing a growing share of them to automation," he says. An oft-cited example is that of Instagram, a photo-sharing app. When it was bought by Facebook in 2012 for $1 billion, it had tens of millions of users, but only 13 employees. Kodak, which once employed 145,000 people making photographic products, went into bankruptcy at around the same time. But such comparisons are misleading, says Marc Andreessen. It was smartphones, not Instagram, that undermined Kodak, and far more people are employed by the smartphone industry and its surrounding ecosystems than ever worked for Kodak or the traditional photography industry.

Is this time different?

So who is right: the pessimists (many of them techie types), who say this time is different and machines really will take all the jobs, or the optimists (mostly economists and historians), who insist that in the end technology always creates more jobs than it destroys? The truth probably lies somewhere in between. AI will not cause mass unemployment, but it will speed up the

existing trend of computer-related automation, disrupting labour markets just as technological change has done before, and requiring workers to learn new skills more quickly than in the past. Mr Bessen predicts a “difficult transition” rather than a “sharp break with history”. But despite the wide range of views expressed, pretty much everyone agrees on the prescription: that companies and governments will need to make it easier for workers to acquire new skills and switch jobs as needed. That would provide the best defence in the event that the pessimists are right and the impact of artificial intelligence proves to be more rapid and more dramatic than the optimists expect.

GEOPOLITICS

The Real Threat of Artificial Intelligence

Kai-Fu Lee, *The New York Times,* June 24, 2017.
https://www.nytimes.com/2017/06/24/opinion/sunday/artificial-intelligence-economic-inequality.html

Editors' Note:
In this article, the author lays out the case for AI being a cause of economic inequality, as it will be difficult for lower skill workers to retrain and obtain jobs that require vastly different skills. The author then goes on to explain how this economic pressure can cause inequality among countries, as those that are advanced in AI technology will be able to redistribute the efficiency gains among their population, while countries that don't have this high level of AI capability will struggle to compete. The author explains how smaller countries that are not the US or China might have difficulty developing their AI capabilities in a way that allows them to match the efficiency gains of countries that are already ahead in AI. The economic inequality caused by AI will take place both within a country, as certain workers are unable to find work, and between countries, as some countries fall behind in AI technology. This article considers the implication of this inequality between countries and the strain that it will put on geopolitical relationships in the future.
~

What worries you about the coming world of artificial intelligence?

Too often the answer to this question resembles the plot of a sci-fi thriller. People worry that developments in A.I. will bring about the "singularity" — that point in history when A.I. surpasses human intelligence, leading to an unimaginable revolution in human affairs. Or they wonder whether instead of our controlling artificial intelligence, it will control us, turning us, in effect, into cyborgs.

These are interesting issues to contemplate, but they are not pressing. They concern situations that may not arise for hundreds of years, if ever. At the moment, there is no known path from our best A.I. tools (like the Google computer program that recently beat the world's best player of the game of Go) to "general" A.I. — self-aware computer programs that can engage in common-sense reasoning, attain knowledge in multiple domains, feel, express and understand emotions and so on.

This doesn't mean we have nothing to worry about. On the contrary, the A.I. products that now exist are improving faster than most people realize and promise to radically transform our world, not always for the better. They are only tools, not a competing form of intelligence. But they will

reshape what work means and how wealth is created, leading to unprecedented economic inequalities and even altering the global balance of power.

It is imperative that we turn our attention to these imminent challenges.

What is artificial intelligence today? Roughly speaking, it's technology that takes in huge amounts of information from a specific domain (say, loan repayment histories) and uses it to make a decision in a specific case (whether to give an individual a loan) in the service of a specified goal (maximizing profits for the lender). Think of a spreadsheet on steroids, trained on big data. These tools can outperform human beings at a given task.

This kind of A.I. is spreading to thousands of domains (not just loans), and as it does, it will eliminate many jobs. Bank tellers, customer service representatives, telemarketers, stock and bond traders, even paralegals and radiologists will gradually be replaced by such software. Over time this technology will come to control semiautonomous and autonomous hardware like self-driving cars and robots, displacing factory workers, construction workers, drivers, delivery workers and many others.

Unlike the Industrial Revolution and the computer revolution, the A.I. revolution is not taking certain jobs (artisans, personal assistants who use paper and typewriters) and replacing them with other jobs (assembly-line workers, personal assistants conversant with computers). Instead, it is poised to bring about a wide-scale decimation of jobs — mostly lower-paying jobs, but some higher-paying ones, too.

This transformation will result in enormous profits for the companies that develop A.I., as well as for the companies that adopt it. Imagine how much money a company like Uber would make if it used only robot drivers. Imagine the profits if Apple could manufacture its products without human labor. Imagine the gains to a loan company that could issue 30 million loans a year with virtually no human involvement. (As it happens, my venture capital firm has invested in just such a loan company.)

We are thus facing two developments that do not sit easily together: enormous wealth concentrated in relatively few hands and enormous numbers of people out of work. What is to be done?

Part of the answer will involve educating or retraining people in tasks A.I. tools aren't good at. Artificial intelligence is poorly suited for jobs involving creativity, planning and "cross-domain" thinking — for example, the work of a trial lawyer. But these skills are typically required by high-paying jobs that may be hard to retrain displaced workers to do. More promising are lower-paying jobs involving the "people skills" that A.I. lacks: social workers, bartenders, concierges — professions requiring nuanced human interaction. But here, too, there is a problem: How many bartenders does a society really need?

The solution to the problem of mass unemployment, I suspect, will involve "service jobs of love." These are jobs that A.I. cannot do, that society needs and that give people a sense of purpose. Examples include accompanying an older person to visit a doctor, mentoring at an orphanage and serving as a sponsor at Alcoholics Anonymous — or, potentially soon, Virtual Reality Anonymous (for those addicted to their parallel lives in computer-generated simulations). The volunteer service jobs of today, in other words, may turn into the real jobs of the future.

Other volunteer jobs may be higher-paying and professional, such as compassionate medical service providers who serve as the "human interface" for A.I. programs that diagnose cancer. In all cases, people will be able to choose to work fewer hours than they do now.

Who will pay for these jobs? Here is where the enormous wealth concentrated in relatively few hands comes in. It strikes me as unavoidable that large chunks of the money created by A.I. will have to be transferred to those whose jobs have been displaced. This seems feasible only through Keynesian policies of increased government spending, presumably raised through taxation on wealthy companies.

As for what form that social welfare would take, I would argue for a conditional universal basic income: welfare offered to those who have a financial need, on the condition they either show an effort to receive training that would make them employable or commit to a certain number of hours of "service of love" voluntarism.

To fund this, tax rates will have to be high. The government will not only have to subsidize most people's lives and work; it will also have to compensate for the loss of individual tax revenue previously collected from employed individuals.

This leads to the final and perhaps most consequential challenge of A.I. The Keynesian approach I have sketched out may be feasible in the United States and China, which will have enough successful A.I. businesses to fund welfare initiatives via taxes. But what about other countries?

They face two insurmountable problems. First, most of the money being made from artificial intelligence will go to the United States and China. A.I. is an industry in which strength begets strength: The more data you have, the better your product; the better your product, the more data you can collect; the more data you can collect, the more talent you can attract; the more talent you can attract, the better your product. It's a virtuous circle, and the United States and China have already amassed the talent, market share and data to set it in motion.

For example, the Chinese speech-recognition company iFlytek and several Chinese face-recognition companies such as Megvii and SenseTime have become industry leaders, as measured by market capitalization. The United States is spearheading the development of autonomous vehicles, led by companies like Google, Tesla and Uber. As for the consumer internet market, seven American or Chinese companies — Google, Facebook, Microsoft, Amazon, Baidu, Alibaba and Tencent — are making extensive use of A.I. and expanding

operations to other countries, essentially owning those A.I. markets. It seems American businesses will dominate in developed markets and some developing markets, while Chinese companies will win in most developing markets.

The other challenge for many countries that are not China or the United States is that their populations are increasing, especially in the developing world. While a large, growing population can be an economic asset (as in China and India in recent decades), in the age of A.I. it will be an economic liability because it will comprise mostly displaced workers, not productive ones.

So if most countries will not be able to tax ultra-profitable A.I. companies to subsidize their workers, what options will they have? I foresee only one: Unless they wish to plunge their people into poverty, they will be forced to negotiate with whichever country supplies most of their A.I. software — China or the United States — to essentially become that country's economic dependent, taking in welfare subsidies in exchange for letting the "parent" nation's A.I. companies continue to profit from the dependent country's users. Such economic arrangements would reshape today's geopolitical alliances.

One way or another, we are going to have to start thinking about how to minimize the looming A.I.-fueled gap between the haves and the have-nots, both within and between nations. Or to put the matter more optimistically: A.I. is presenting us with an opportunity to rethink economic inequality on a global scale. These challenges are too far-ranging in their effects for any nation to isolate itself from the rest of the world.

China's Plan to 'Lead' in AI: Purpose, Prospects, and Problems

Graham Webster, Rogier Creemers, Paul Triolo, and Elsa Kania, *New America Foundation*, August 1, 2017.
https://www.newamerica.org/cybersecurity-initiative/blog/chinas-plan-lead-ai-purpose-prospects-and-problems/

Editors' Note:
The New America Foundation is a liberal Washington-based think tank that focuses on U.S. national security policy. This paper analyzes the Chinese State Council's "New Generation Artificial Intelligence Development Plan," which was released on July 20, 2017. It explains how the Chinese government intends to make China the international leader in A.I. research—and what this means for Chinese politics, defense, and industry.
~

The present global verve about artificial intelligence (AI) and machine learning technologies has resonated in China as much as anywhere on earth. With the State Council's issuance of the "New Generation Artificial Intelligence Development Plan" (新一代人工智能发展规划) on July 20, China's government set out an ambitious roadmap including targets through 2030. Meanwhile, in China's leading cities, flashy conferences on AI have become commonplace. It seems every mid-sized tech company wants to show off its self-driving car efforts, while numerous financial tech start-ups tout an AI-driven approach. Chatbot startups clog investors' date books, and Shanghai metro ads pitch AI-taught English language learning.

The surge of showmanship and investment in the private sector is paralleled by a new focus among officials and policy thinkers in academia, civilian and military sectors of government, and major companies. Questions about how to regulate AI, and how to develop and use it ethically, have become a major topic for China's digital policy brain trust in recent months.

At this point, investors, policymakers, and even engineers might wonder where the hype ends and the reality of China's agenda for "AI 2.0" begins. This new development plan is certainly (and typically for an aspirational central government document) packed with vagaries and grandiose ambitions, as China declares its intentions to pursue a "first-mover advantage" to become "the world's primary AI innovation center" by 2030. Nonetheless, the plan contains real signals and measures worthy of attention. Its specific and nonspecific goals, its bureaucratic positioning, and its long time-horizon make it an important reference point for a wide variety of policy, business, and security developments in coming years.

In laying out its top-line goals for economic development and AI, the new State Council plan declares that in under a decade, AI will become "the main driving force for China's industrial upgrading and economic transformation." Statements such as these underline the ambition

captured in the plan itself, but as always in Chinese politics, attention is also due to whose interests and ambitions are driving a given agenda.

Promising government investment, and recognizing where China lags. The plan prescribes a high level of government investment in theoretical and applied AI breakthroughs, while also acknowledging that, in China as around the world, private companies are currently leading the charge on commercial applications of AI. Large companies in cloud services, e-commerce, social media, or other sectors with access to large troves of data that can be used to train AI algorithms are naturally positioned to lead in a variety of fields, including facial recognition, voice recognition, and natural language processing. The plan acknowledges, meanwhile, that China remains far behind world leaders in development of key hardware enablers of AI, such as microchips suited for machine learning use (e.g., GPUs or re-confirgurable processors). The plan's ambition is underlined by its recognition of the hard road ahead.

Reconciling advances in AI with risks of disruption. With the proliferation of AI, China's government recognizes that new risks and challenges will arise for governance, economic security, and social stability. The plan also focuses on minimizing these risks to ensure the "safe, reliable, and controllable" development of AI. The plan includes formulation of laws, regulations, and ethical norms on AI, as well as mechanisms for safety and supervision. The plan seeks to mitigate likely negative externalities, such as job losses, associated with AI, while fully leveraging the opportunities.

The Political Layer

Harnessing the rise of AI to keep science and technology bureaucrats relevant. The new plan is clearly driven by China's waning Ministry of Science and Technology (MOST) and related science and technology (S&T) bureaucracy, as it links progress in AI with virtually every other S&T program of note. In the plan, the a National S&T Structural Reform and Innovation Systems Construction Leading Small Group, one of the several coordinating bodies (including the sometimes competing "Cybersecurity and Informatization" group that sits above the Cyberspace Administration of China) through which President Xi Jinping exercises influence and power, is heralded as the driver of the AI program and particularly the AI-related legal and regulatory system that the plan calls for. In addition, the plan calls to create a new AI Plan Promotion Office within MOST.

For China's government, another digital solution to provide public goods. One theme of the plan is that AI can serve as a vehicle through which the Chinese government can provide better governance to the Chinese people, using AI to drive smart cities, smart government, smart manufacturing, and forming the infrastructure for a smart society. According to the plan's lofty aspirations, AI applications in agriculture, transportation, social security, pension management, public security, and a host of other government functions will enable the government to provide new levels of service and benefit to the Chinese nation. Informatization-linked governance improvements are not a new idea for China. A decade ago, governance sections in big Chinese

bookstores started to fill with future-oriented e-government titles, which are joined now by guides on modernizing government services through the earlier official "Internet Plus" concept, and now how to interact with the public through microblogs and Tencent's ubiquitous WeChat platform.

Meanings of Leadership in AI

Government scientists, S&T bureaucrats, and planners are unlikely to substantively lead China's AI development when private companies are the ones developing crucial data resources and are much more able to attract and pay for top AI talent. Existing momentum in the private sector will already doubtless make Chinese efforts among the world leaders in several types of AI applications by 2030, though it may not be a government plan that makes the difference.

The question remains what that leadership means, as many AI applications are developed based on culturally, linguistically, and geographically bounded data. For instance, if a Chinese company develops natural language processing technology that gives Chinese users a level of capability unmatched globally, that doesn't mean it can necessarily market the same service in other languages. Other more mundane challenges include ownership of high resolution digital maps that will be required for autonomous driving. At very least, the notion of one nation leading in AI generally will be complicated by the field's diversity of technical and social challenges, as well as the different inputs necessary for different applications—not to mention that a comprehensively leading effort will by definition take place across borders.

The plan's recognition of the need for regulatory, legal, and ethical principles for AI development and use does, however, represent an uncommonly foresighted approach. Of course, the Chinese government's approach to AI regulation, ethics, and economic adjustment will reflect is broader model of governance and ideology. Thus it will be crucial for other jurisdictions, for instance the United States and the EU, to develop regulatory, ethical, and developmental approaches that reflect their own values.

BIAS, AI, AND SOCIETY

AI programs exhibit racial and gender biases, research reveals

Hannah Devlin, *The Guardian,* April 13, 2017
https://www.theguardian.com/technology/2017/apr/13/ai-programs-exhibit-racist-and-sexist-biases-research-reveals

Editors' Note:
In this article, it is noted that AI systems are representative of the data that we put into them, and that if this data is flawed or biased then the outputs of the AI system will be as well. Sexism or racism that exists in the current state, then becomes part of a system that has identified that pattern and continues to repeat it. This is something that those who "teach" AI systems need to guard against so that these biases hopefully do not continue to be perpetuated by AI. The article explores these issues mostly by looking at current developments in computer learning of language.
~

An artificial intelligence tool that has revolutionised the ability of computers to interpret everyday language has been shown to exhibit striking gender and racial biases.

The findings raise the spectre of existing social inequalities and prejudices being reinforced in new and unpredictable ways as an increasing number of decisions affecting our everyday lives are ceded to automatons.

In the past few years, the ability of programs such as Google Translate to interpret language has improved dramatically. These gains have been thanks to new machine learning techniques and the availability of vast amounts of online text data, on which the algorithms can be trained.

However, as machines are getting closer to acquiring human-like language abilities, they are also absorbing the deeply ingrained biases concealed within the patterns of language use, the latest research reveals.

Joanna Bryson, a computer scientist at the University of Bath and a co-author, said: "A lot of people are saying this is showing that AI is prejudiced. No. This is showing we're prejudiced and that AI is learning it."

But Bryson warned that AI has the potential to reinforce existing biases because, unlike humans, algorithms may be unequipped to consciously counteract learned biases. "A danger would be if you had an AI system that didn't have an explicit part that was driven by moral ideas, that would be bad," she said.

The research, published in the journal Science, focuses on a machine learning tool known as "word embedding", which is already transforming the way computers interpret speech and text.

Some argue that the natural next step for the technology may involve machines developing human-like abilities such as common sense and logic.

“A major reason we chose to study word embeddings is that they have been spectacularly successful in the last few years in helping computers make sense of language,” said Arvind Narayanan, a computer scientist at Princeton University and the paper’s senior author.

The approach, which is already used in web search and machine translation, works by building up a mathematical representation of language, in which the meaning of a word is distilled into a series of numbers (known as a word vector) based on which other words most frequently appear alongside it. Perhaps surprisingly, this purely statistical approach appears to capture the rich cultural and social context of what a word means in the way that a dictionary definition would be incapable of.

For instance, in the mathematical “language space”, words for flowers are clustered closer to words linked to pleasantness, while words for insects are closer to words linked to unpleasantness, reflecting common views on the relative merits of insects versus flowers.

The latest paper shows that some more troubling implicit biases seen in human psychology experiments are also readily acquired by algorithms. The words “female” and “woman” were more closely associated with arts and humanities occupations and with the home, while “male” and “man” were closer to maths and engineering professions.

And the AI system was more likely to associate European American names with pleasant words such as “gift” or “happy”, while African American names were more commonly associated with unpleasant words.

The findings suggest that algorithms have acquired the same biases that lead people (in the UK and US, at least) to match pleasant words and white faces in implicit association tests.

These biases can have a profound impact on human behaviour. One previous study showed that an identical CV is 50% more likely to result in an interview invitation if the candidate’s name is European American than if it is African American. The latest results suggest that algorithms, unless explicitly programmed to address this, will be riddled with the same social prejudices.

“If you didn’t believe that there was racism associated with people’s names, this shows it’s there,” said Bryson.

The machine learning tool used in the study was trained on a dataset known as the “common crawl” corpus – a list of 840bn words that have been taken as they appear from material published online. Similar results were found when the same tools were trained on data from Google News.

Sandra Wachter, a researcher in data ethics and algorithms at the University of Oxford, said: "The world is biased, the historical data is biased, hence it is not surprising that we receive biased results."

Rather than algorithms representing a threat, they could present an opportunity to address bias and counteract it where appropriate, she added.

"At least with algorithms, we can potentially know when the algorithm is biased," she said. "Humans, for example, could lie about the reasons they did not hire someone. In contrast, we do not expect algorithms to lie or deceive us."

However, Wachter said the question of how to eliminate inappropriate bias from algorithms designed to understand language, without stripping away their powers of interpretation, would be challenging.

"We can, in principle, build systems that detect biased decision-making, and then act on it," said Wachter, who along with others has called for an AI watchdog to be established. "This is a very complicated task, but it is a responsibility that we as society should not shy away from."

MACROECONOMICS

Why Artificial Intelligence is the Future of Growth

Mark Purdy and Paul Daugherty, *Accenture,* 2016.
https://www.accenture.com/t20170206T005353Z__w__/us-en/_acnmedia/PDF-33/Accenture-Why-AI-is-the-Future-of-Growth.PDFla=en#zoom=50

Editors' Note:
This article is from Accenture, a large strategy and consulting firm that focuses on how new technologies will affect established businesses. In this report, they discuss the problems with the current economic state, such as stagnant growth, and how the efficiencies created through increased use in AI can actually cause the economies of developed nations to grow much faster. The report explains how rather than causing many problems to the current economic state, AI will actually improve output for most countries. Especially useful is the discussion of the diffusion of the benefits of AI through the economy, using the example of driverless vehicles. Also take note of the specific examples of applications of AI.
~

There has been marked decline in the ability of increases in capital investment and in labor to propel economic progress. These two levers are the traditional drivers of production, yet they are no longer able to sustain the steady march of prosperity enjoyed in previous decades in most developed economies. But long-term pessimism is unwarranted. With the recent convergence of a transformative set of technologies, economies are entering a new era in which artificial intelligence (AI) has the potential to overcome the physical limitations of capital and labor and open up new sources of value and growth. Increases in capital and labor are no longer driving the levels of economic growth the world has become accustomed to and desires.

Fortunately, a new factor of production is on the horizon, and it promises to transform the basis of economic growth for countries across the world. Indeed, Accenture analyzed 12 developed economies and found that AI has the potential to double their annual economic growth rates by 2035.

[...]

Today, we are witnessing the takeoff of another transformative set of technologies, commonly referred to as artificial intelligence […] Many see AI as similar to past technological inventions. If we believe this, then we can expect some growth, but nothing transformational. But what if AI has the potential to be not just another driver of TFP, but an entirely new factor of production? How can this be?

The key is to see AI as a capital-labor hybrid. AI can replicate labor activities at much greater scale and speed, and to even perform some tasks beyond the capabilities of humans. Not to

mention that in some areas it has the ability to learn faster than humans, if not yet as deeply. For example, by using virtual assistants, 1,000 legal documents can be reviewed in a matter of days instead of taking three people six months to complete.
Similarly, AI can take the form of physical capital such as robots and intelligent machines. And unlike conventional capital, such as machines and buildings, it can actually improve over time, thanks to its self-learning capabilities.

[...]

The new AI-powered wave of intelligent automation is already creating growth through a set of features unlike those of traditional automation solutions. The first feature is its ability to automate complex physical world tasks that require adaptability and agility. Consider the work of retrieving items in a warehouse, where companies have relied on people's ability to navigate crowded spaces and avoid moving obstacles. Now, robots from Fetch Robotics use lasers and 3D depth-sensors to navigate safely and work alongside warehouse workers. Used in tandem with people, the robots can handle the vast majority of items in a typical warehouse.

Whereas traditional automation technology is task specific, the second distinct feature of AI-powered intelligent automation is its ability to solve problems across industries and job titles. For instance, Amelia—an AI platform by IPsoft with natural language processing capabilities—has supported maintenance engineers in remote locations. Having read all the manuals, Amelia can diagnose a problem and suggest a solution. This platform has also learned the answers to the 120 questions most frequently asked by mortgage brokers and has been used in a bank to handle such financial queries, traditionally a labor intensive task.

The third and most powerful feature of intelligent automation is self-learning, enabled by repeatability at scale. Amelia, like a conscientious employee, recognizes the gaps in her own knowledge and takes steps to close them. If Amelia is presented with a question that she cannot answer, she escalates it to a human colleague, then observes how the person solves the problem. The self-learning aspect of AI is a fundamental change. Whereas traditional automation capital degrades over time, intelligent automation assets constantly improve.

A significant part of the economic growth from AI will come not from replacing existing labor and capital, but in enabling them to be used much more effectively. For example, AI can enable humans to focus on parts of their role that add the most value. Hotel staff spend a lot of their time making routine room deliveries. Why not assign the task to Relay, an autonomous service industry robot developed by Savioke, instead? Last year, the Relay fleet made more than 11,000 guest deliveries in the five large hotel chains where it is deployed. As Steve Cousins, CEO of Savioke, told us: "Relay enables staff to redirect their time toward increasing customer satisfaction."

Also, AI augments labor by complementing human capabilities, offering employees new tools to enhance their natural intelligence. For example, Praedicat, a company providing risk modeling services to property and casualty insurers, is improving underwriters' risk-pricing abilities. Using machine learning and big data processing technologies, its AI platform reads more than 22

million peer reviewed scientific papers to identify serious emerging risks. As a result, underwriters can not only price risk more accurately, but also create new insurance products.

AI can also improve capital efficiency— a crucial factor in industries where it represents a large sunk cost. For instance, in manufacturing, industrial robotics company Fanuc has teamed up with Cisco and other firms to create a platform to reduce factory downtime—estimated at one major automotive manufacturer to cost US$20,000 per minute.

The Fanuc Intelligent Edge Link and Drive (FIELD) system is an analytics platform powered by advanced machine learning. It captures and analyzes data from disparate parts of the manufacturing process to improve manufacturing production. Already FIELD has been deployed in an 18-month "zero downtime" trial at one manufacturer, where it realized significant cost savings.

Innovation diffusion
One of the least-discussed benefits of artificial intelligence is its ability to propel innovations as it diffuses through the economy.

Take driverless vehicles. Using a combination of lasers, global positioning systems, radar, cameras, computer vision and machine learning algorithms, driverless vehicles can enable a machine to sense its surroundings and act accordingly. Not only are Silicon Valley technology companies entering the market, but traditional companies are building new partnerships to stay relevant.

The insurance industry could create new revenue streams from the masses of data that self-driving vehicles generate. By combining vehicle data with other streams such as smart-phones and public transport systems, they could not only build up a more complete picture of their customers, but also they could create new policies that insure total customer mobility, not just driving.

Real-time, accurate road and traffic data generated by driverless vehicles could supplement other sources of information to enable local authorities to change the way they charge for road usage. Standard vehicle registration could be replaced with more equitable and convenient pay-per-use road tolls, with instantly updated prices to help reduce congestion. [...]

As innovation begets innovation, the potential impact of driverless vehicles on economies could eventually extend well beyond the automotive industry. Mobile service providers could see even more demand from subscribers as drivers, now free to enjoy leisure activities while traveling, spend more time on the Internet, which, in turn, could create new advertising opportunities for the service providers and selling opportunities for their retailer partners.

There could even be significant social benefits. Driverless vehicles are expected to reduce the number of road accidents and traffic fatalities dramatically, making the technology potentially one of the most transformative public health initiatives in human history. They could also give back independence to people who cannot drive due to disability, enabling them to take up jobs

from which they were previously excluded. And, even among those who can drive, driverless cars will make traveling far more convenient, freeing up time that people can dedicate to work or leisure.

[…]

Many commentators are concerned that AI will eliminate jobs, worsen inequality and erode incomes. This explains the rise in protests around the world and discussions taking place in countries, such as Switzerland, on the introduction of a universal basic income. Policy makers must recognize that these apprehensions are valid.

Their response should be twofold. First, policy makers should highlight how AI can result in tangible benefits. For instance, AI can improve job satisfaction. An Accenture survey highlighted that 84 percent of managers believe machines will make them more effective and their work more interesting.

Beyond the workplace, AI promises to alleviate some of the world's greatest problems, such as climate change (through more efficient transportation) and poor access to healthcare (by reducing the strain on overloaded systems). Benefits like these should be clearly articulated to encourage a more positive outlook on AI's potential

Second, policy makers need to actively address and preempt the downsides of AI. Some groups will be affected disproportionately by these changes. To prevent a backlash, policy makers should identify the groups at high risk of displacement and create strategies that focus on reintegrating them into the economy.

Artificial Intelligence, Automation, and the Economy

Jason Furman, John P. Holdren et al. *The Executive Office of the President*, 2016. https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF

Editors' Note:
The following is a summary of a report from the Obama White House in 2016 considering the possible effects of AI on the US economy and how the possible issues of inequality and unemployment can be dealt with through policy. It is important to note the three suggested policy responses to increasing automation.
~

Accelerating artificial intelligence (AI) capabilities will enable automation of some tasks that have long required human labor.1 These transformations will open up new opportunities for individuals, the economy, and society, but they have the potential to disrupt the current livelihoods of millions of Americans. Whether AI leads to unemployment and increases in inequality over the long-run depends not only on the technology itself but also on the institutions and policies that are in place. This report examines the expected impact of AI-driven automation on the economy, and describes broad strategies that could increase the benefits of AI and mitigate its costs.

Economics of AI-Driven Automation

Technological progress is the main driver of growth of GDP per capita, allowing output to increase faster than labor and capital. One of the main ways that technology increases productivity is by decreasing the number of labor hours needed to create a unit of output. Labor productivity increases generally translate into increases in average wages, giving workers the opportunity to cut back on work hours and to afford more goods and services. Living standards and leisure hours could both increase, although to the degree that inequality increases—as it has in recent decades—it offsets some of those gains.

AI should be welcomed for its potential economic benefits. Those economic benefits, however, will not necessarily be evenly distributed across society. For example, the 19th century was characterized by technological change that raised the productivity of lower-skilled workers relative to that of higher-skilled workers. Highly-skilled artisans who controlled and executed full production processes saw their livelihoods threatened by the rise of mass production technologies. Ultimately, many skilled crafts were replaced by the combination of machines and lower-skilled labor. Output per hour rose while inequality declined, driving up average living standards, but the labor of some high-skill workers was no longer as valuable in the market.

In contrast, technological change tended to work in a different direction throughout the late 20th century. The advent of computers and the Internet raised the relative productivity of higher skilled workers. Routine-intensive occupations that focused on predictable, easily-programmable tasks—such as switchboard operators, filing clerks, travel agents, and assembly line workers—were particularly vulnerable to replacement by new technologies. Some occupations were virtually eliminated and demand for others reduced. Research suggests that technological innovation over this period increased the productivity of those engaged in abstract thinking, creative tasks, and problem-solving and was therefore at least partially responsible for the substantial growth in jobs employing such traits. Shifting demand towards more skilled labor raised the relative pay of this group, contributing to rising inequality. At the same time, a

slowdown in the rate of improvement in education, and institutional changes such as the reduction in unionization and decline in the minimum wage, also contributed to inequality—underscoring that technological changes do not uniquely determine outcomes.

Today, it may be challenging to predict exactly which jobs will be most immediately affected by AI-driven automation. Because AI is not a single technology, but rather a collection of technologies that are applied to specific tasks, the effects of AI will be felt unevenly through the economy. Some tasks will be more easily automated than others, and some jobs will be affected more than others—both negatively and positively. Some jobs may be automated away, while for others, AI-driven automation will make many workers more productive and increase demand for certain skills. Finally, new jobs are likely to be directly created in areas such as the development and supervision of AI as well as indirectly created in a range of areas throughout the economy as higher incomes lead to expanded demand.

Recent research suggests that the effects of AI on the labor market in the near term will continue the trend that computerization and communication innovations have driven in recent decades. Researchers' estimates on the scale of threatened jobs over the next decade or two range from 9 to 47 percent. For context, every 3 months about 6 percent of jobs in the economy are destroyed by shrinking or closing businesses, while a slightly larger percentage of jobs are added—resulting in rising employment and a roughly constant unemployment rate. The economy has repeatedly proven itself capable of handling this scale of change, although it would depend on how rapidly the changes happen and how concentrated the losses are in specific occupations that are hard to shift from.

Research consistently finds that the jobs that are threatened by automation are highly concentrated among lower-paid, lower-skilled, and less-educated workers. This means that automation will continue to put downward pressure on demand for this group, putting downward pressure on wages and upward pressure on inequality. In the longer-run, there may be different or larger effects. One possibility is superstar-biased technological change, where the benefits of technology accrue to an even smaller portion of society than just highly-skilled workers. The winner-take-most nature of information technology markets means that only a few may come to dominate markets. If labor productivity increases do not translate into wage increases, then the large economic gains brought about by AI could accrue to a select few. Instead of broadly shared prosperity for workers and consumers, this might push towards reduced competition and increased wealth inequality.

Historically and across countries, however, there has been a strong relationship between productivity and wages—and with more AI the most plausible outcome will be a combination of higher wages and more opportunities for leisure for a wide range of workers. But the degree that this materializes depends not just on the nature of technological change but importantly on the policy and institutional choices that are made about how to prepare workers for AI and to handle its impacts on the labor market.

Policy Responses

Technology is not destiny; economic incentives and public policy can play a significant role in shaping the direction and effects of technological change. Given appropriate attention and the right policy and institutional responses, advanced automation can be compatible with productivity, high levels of employment, and more broadly shared prosperity. In the past, the U.S. economy has adapted to new production patterns and maintained high levels of employment alongside rising productivity as more productive workers have had more incentive to work and more highly paid workers have spent more, supporting this work. But, some shocks have left a growing share of workers out of the labor force. This report advocates strategies to educate and prepare new workers to enter the workforce, cushion workers who lose jobs, keep them attached to the labor force, and combat inequality. Most of these strategies would be important regardless of AI-driven automation, but all take on even greater importance to the degree that AI is making major changes to the economy.

Strategy #1: Invest in and develop AI for its many benefits.
If care is taken to responsibly maximize its development, AI will make important, positive contributions to aggregate productivity growth, and advances in AI technology hold incredible potential to help the United States stay on the cutting edge of innovation. Government has an important role to play in advancing the AI field by investing in research and development. Among the areas for advancement in AI are cyber defense and the detection of fraudulent transactions and messages. In addition, the rapid growth of AI has also dramatically increased the need for people with relevant skills from all backgrounds to support and advance the field. Prioritizing diversity and inclusion in STEM fields and in the AI community specifically, in addition to other possible policy responses, is a key part in addressing potential barriers stemming from algorithmic bias. Competition from new and existing firms, and the development of sound pro-competition policies, will increasingly play an important role in the creation and adoption of new technologies and innovations related to AI.

Strategy #2: Educate and train Americans for jobs of the future.
As AI changes the nature of work and the skills demanded by the labor market, American workers will need to be prepared with the education and training that can help them continue to succeed. Delivering this education and training will require significant investments. This starts with providing all children with access to high-quality early education so that all families can prepare their students for continued education, as well as investing in graduating all students from high school college- and career ready, and ensuring that all Americans have access to affordable post-secondary education. Assisting U.S. workers in successfully navigating job transitions will also become increasingly important; this includes expanding the availability of job-driven training and opportunities for lifelong learning, as well as providing workers with improved guidance to navigate job transitions.

Strategy #3: Aid workers in the transition and empower workers to ensure broadly shared growth.
Policymakers should ensure that workers and job seekers are both able to pursue the job opportunities for which they are best qualified and best positioned to ensure they receive an

appropriate return for their work in the form of rising wages. This includes steps to modernize the social safety net, including exploring strengthening critical supports such as unemployment insurance, Medicaid, Supplemental Nutrition Assistance Program (SNAP), and Temporary Assistance for Needy Families (TANF), and putting in place new programs such as wage insurance and emergency aid for families in crisis. Worker empowerment also includes bolstering critical safeguards for workers and families in need, building a 21st century retirement system, and expanding healthcare access. Increasing wages, competition, and worker bargaining power, as well as modernizing tax policy and pursuing strategies to address differential geographic impact, will be important aspects of supporting workers and addressing concerns related to displacement amid shifts in the labor market.

Finally, if a significant proportion of Americans are affected in the short- and medium-term by AI-driven job displacements, policymakers will need to consider more robust interventions, such as further strengthening the unemployment insurance system and countervailing job creation strategies, to smooth the transition.

Conclusion
Responding to the economic effects of AI-driven automation will be a significant policy challenge for the next Administration and its successors. AI has already begun to transform the American workplace, change the types of jobs available, and reshape the skills that workers need in order to thrive. All Americans should have the opportunity to participate in addressing these challenges, whether as students, workers, managers, technical leaders, or simply as citizens with a voice in the policy debate.

AI raises many new policy questions, which should be continued topics for discussion and consideration by future Administrations, Congress, the private sector, academia, and the public. Continued engagement among government, industry, technical and policy experts, and the public should play an important role in moving the Nation toward policies that create broadly shared prosperity, unlock the creative potential of American companies and workers, and ensure America’s continued leadership in the creation and use of AI.

PRIVACY CONCERNS

Artificial intelligence is ripe for abuse, tech researcher warns: 'a fascist's dream'

Olivia Solon, *The Guardian, March 13, 2017.*
https://www.theguardian.com/technology/2017/mar/13/artificial-intelligence-ai-abuses-fascism-donald-trump

Editors' Note:
In this article, a security researcher describes some possible totalitarian style abuses that would be possible utilizing advanced AI systems. The issue is that surveillance already occurs, but that it could occur with a scale and specificity unlike what happens today.

For more information about possible invasions of privacy utilizing AI, see the section below on existential risks, especially the first article on why humans are the reason AI is scary.
~

As artificial intelligence becomes more powerful, people need to make sure it's not used by authoritarian regimes to centralize power and target certain populations, Microsoft Research's Kate Crawford warned on Sunday.

In her SXSW session, titled Dark Days: AI and the Rise of Fascism, Crawford, who studies the social impact of machine learning and large-scale data systems, explained ways that automated systems and their encoded biases can be misused, particularly when they fall into the wrong hands.

"Just as we are seeing a step function increase in the spread of AI, something else is happening: the rise of ultra-nationalism, rightwing authoritarianism and fascism," she said.

All of these movements have shared characteristics, including the desire to centralize power, track populations, demonize outsiders and claim authority and neutrality without being accountable. Machine intelligence can be a powerful part of the power playbook, she said.

One of the key problems with artificial intelligence is that it is often invisibly coded with human biases. She described a controversial piece of research from Shanghai Jiao Tong University in China, where authors claimed to have developed a system that could predict criminality based on someone's facial features. The machine was trained on Chinese government ID photos, analyzing the faces of criminals and non-criminals to identify predictive features. The researchers claimed it was free from bias.

"We should always be suspicious when machine learning systems are described as free from bias if it's been trained on human-generated data," Crawford said. "Our biases are built into that training data."

In the Chinese research it turned out that the faces of criminals were more unusual than those of law-abiding citizens. "People who had dissimilar faces were more likely to be seen as untrustworthy by police and judges. That's encoding bias," Crawford said. "This would be a terrifying system for an autocrat to get his hand on."

Crawford then outlined the "nasty history" of people using facial features to "justify the unjustifiable". The principles of phrenology, a pseudoscience that developed across Europe and

the US in the 19th century, were used as part of the justification of both slavery and the Nazi persecution of Jews.

With AI this type of discrimination can be masked in a black box of algorithms, as appears to be the case with a company called Faceception, for instance, a firm that promises to profile people's personalities based on their faces. In its own marketing material, the company suggests that Middle Eastern-looking people with beards are "terrorists", while white looking women with trendy haircuts are "brand promoters".

Another area where AI can be misused is in building registries, which can then be used to target certain population groups. Crawford noted historical cases of registry abuse, including IBM's role in enabling Nazi Germany to track Jewish, Roma and other ethnic groups with the Hollerith Machine, and the Book of Life used in South Africa during apartheid.

Donald Trump has floated the idea of creating a Muslim registry. "We already have that. Facebook has become the default Muslim registry of the world," Crawford said, mentioning research from Cambridge University that showed it is possible to predict people's religious beliefs based on what they "like" on the social network. Christians and Muslims were correctly classified in 82% of cases, and similar results were achieved for Democrats and Republicans (85%). That study was concluded in 2013, since when AI has made huge leaps .

Crawford was concerned about the potential use of AI in predictive policing systems, which already gather the kind of data necessary to train an AI system. Such systems are flawed, as shown by a Rand Corporation study of Chicago's program. The predictive policing did not reduce crime, but did increase harassment of people in "hotspot" areas. Earlier this year the justice department concluded that Chicago's police had for years regularly used "unlawful force", and that black and Hispanic neighborhoods were most affected.

Another worry related to the manipulation of political beliefs or shifting voters, something Facebook and Cambridge Analytica claim they can already do. Crawford was skeptical about giving Cambridge Analytica credit for Brexit and the election of Donald Trump, but thinks what the firm promises – using thousands of data points on people to work out how to manipulate their views – will be possible "in the next few years".

"This is a fascist's dream," she said. "Power without accountability."

Such black box systems are starting to creep into government. Palantir is building an intelligence system to assist Donald Trump in deporting immigrants.

"It's the most powerful engine of mass deportation this country has ever seen," she said.

But what do you do if the system has got something wrong? What if it has incorrect data?

Crawford argues that we have to make these AI systems more transparent and accountable. “The ocean of data is so big. We have to map their complex subterranean and unintended effects.”

Crawford has founded AI Now, a research community focused on the social impacts of artificial intelligence to do just this

“We want to make these systems as ethical as possible and free from unseen biases.”

Amazon’s Echo Look is a minefield of AI and privacy concerns

James Vincent, *The Verge,* April 27, 2017.
https://www.theverge.com/2017/4/27/15447834/amazons-echo-look-ai-analysis-concerns

Computer scientist Andrew Ng once described the power of contemporary AI as the ability to automate any mental task that takes a human “less than one second of thought.” It’s a rule of thumb that’s worth remembering when you think about Amazon’s new Echo Look — a smart camera with a built-in AI assistant. Amazon says the Echo Look will help users dress and give them fashion advice, but what other judgements could it make?

Think about it. If you pass a stranger in the street, how much information do you get with a second's glance? You can probably make some decent estimates about their height, weight, age, race, and gender. If they were far enough along in pregnancy, you'd know. And you could take a stab at other, more contentious questions, like: are they rich or poor? Friendly or closed-off? Are they having a good day? If it takes a second for you to answer these questions, then Amazon's AI could take a stab at it as well. And, given enough data, it can offer far, far more accurate answers.

As academic and sociologist Zeynep Tufekci put it on Twitter: "Machine learning algorithms can do so much with regular full length pictures of you. They can infer private things you did not disclose [...] All this to sell you more clothes. We are selling out to surveillance capitalism that can quickly evolve into authoritarianism for so cheap." (The whole thread from Tufecki is definitely worth a read.)

This might seem overly speculative or alarmist to some, but Amazon isn't offering much reassurance about what they plan to do with data gathered from the Echo Look. representative for the company told The Verge, that at this point, the Look will only use machine learning to analyze users' fashion choices, but when asked if this might change in the future, they said they "can't speculate" on the topic. The rep stressed that users can delete videos and photos taken by the Look at any time, but until they do, it seems this content will be stored indefinitely on Amazon's servers.

This non-denial means the Echo Look could, in the future, provide Amazon with the resource every AI company craves: data. And full-length photos of people taken regularly in the same location would be a particularly valuable dataset — even more so if you combine this information with everything else Amazon knows about its customers (their shopping habits, for one). But when asked whether the company would ever combine these two datasets, an Amazon rep only gave the same, canned answer: "Can't speculate."

The company did, though, say it wouldn't share any personal information gleaned from the Echo Look to "advertisers or to third-party sites that display our interest-based ads." That means Amazon could still use data from the Look to target ads at you itself, but at least third parties won't.

Right now, the Echo Look is halfway between prototype and full-on product. As is often the case with Amazon's hardware efforts, the company seems most interested in just getting a product out there and gauging public reaction, rather than finessing every detail. The company is giving no indication of when the Echo Look will actually be available, and it's currently only being sold "by invitation only." All this means that Amazon itself probably isn't yet sure what exactly it will do with the data the device collects. But, if the company refuses to give any more detail, it's understandable to fear the worst.

EXISTENTIAL RISKS OF AI

Humans, Not Robots, Are the Real Reason Artificial Intelligence Is Scary

Zach Musgrave and Bryan W. Roberts, *The Atlantic Monthly,* August 14, 2015.
https://www.theatlantic.com/technology/archive/2015/08/humans-not-robots-are-the-real-reason-artificial-intelligence-is-scary/400994/

Editors' Note:
The authors of this article explain that once created, autonomous weapons could be a very cost-effective way for small nations, terrorist groups, or others to inflict tremendous harm on civilian populations. They see this as a much larger threat than a possible AI takeover and destruction of mankind. Humans in control of AI aided weapons present a much more realistic threat, especially when one considers the risk of hackers repurposing these systems. This is what discussions on regulating AI should focus on, according to the authors.
~

Unfortunately, much of the recent outcry against artificial-intelligence weapons has been confused, conjuring robot takeovers of mankind. This scenario is implausible in the near term, but AI weapons actually do present a danger not posed by conventional, human-controlled weapons, and there is good reason to ban them.

We've already seen a glimpse of the future of artificial intelligence in Google's self-driving cars. Now imagine that some fiendish crime syndicate were to steal such a car, strap a gun to the top, and reprogram it to shoot people. That's an AI weapon.

The potential of these weapons has not escaped the imaginations of governments. This year we saw the US Navy's announcement of plans to develop autonomous-drone weapons, as well as the announcement of both the South Korean Super aEgis II automatic turret and the Russian Platform-M automatic combat machine.

But governments aren't the only players making AI weapons. Imagine a GoPro-bearing quadcopter drone, the kind of thing anyone can buy. Now imagine a simple piece of software that allows it to fly automatically. The same nefarious crime syndicate that can weaponize a driverless car is just inches away from attaching a gun and programming it to kill people in a crowded public place.

This is the immediate danger with AI weapons: They are easily converted into indiscriminate death machines, far more dangerous than the same weapons with a human at the helm.

Stephen Hawking and Max Tegmark, alongside Elon Musk and many others have all signed a Future of Life petition to ban AI weapons, hosted by the institution that received a $10 million donation from Mr. Musk in January. This followed a UN meeting on 'killer robots' in April that did not lead to any lasting policy decisions. The letter accompanying the Future of Life petition suggests the danger of AI weapons is immediate, requiring action to avoid disasters within the next few years at the earliest. Unfortunately, it doesn't explain what sorts of AI weapons are on the immediate horizon.

Unlike self-aware computer networks, self-driving cars with machine guns are possible right now.

Many have expressed concerns about apocalyptic Terminator-like scenarios, in which robots develop the human-like ability to interact with the world all by themselves and attempt to conquer it. For example, physicist and Astronomer Royal Sir Martin Rees warned of catastrophic scenarios like "dumb robots going rogue or a network that develops a mind of its own." His Cambridge colleague and philosopher Huw Price has voiced a similar concern that humans may not survive when intelligence "escapes the constraints of biology." Together the two helped create the Centre for the Study of Existential Risk at the University of Cambridge to help avoid such dramatic threats to human existence.

These scenarios are certainly worth studying. However, they are far less plausible and far less immediate than the AI-weapons danger on the horizon now.

How close are we to developing the human-like artificial intelligence? By almost all standards, the answer is: not very close. University of Reading chatbot 'Eugene Goostman' was reported by many media outlets to be truly intelligent because it managed to fool a few humans into thinking it was a real 13-year-old boy. However, the chatbot turned out to be miles away from real human-like intelligence, as computer scientist Scott Aaronson demonstrated by destroying Eugene with his first question, "Which is bigger, a shoebox or Mt Everest?" After completely flubbing the answer, and then stumbling on, "How many legs does a camel have?" the emperor was revealed to be without clothes.

In spite of all this, we, the authors of this article, have both signed the Future of Life petition against AI weapons. Here's why: Unlike self-aware computer networks, self-driving cars with machine guns are possible right now. The problem with such AI weapons is not that they are on the verge of taking over the world. The problem is that they are trivially easy to reprogram, allowing anyone to create an efficient and indiscriminate killing machine at an incredibly low cost. The machines themselves aren't what's scary. It's what any two-bit hacker can do with them on a relatively modest budget.

Imagine an up-and-coming despot who would like to eliminate opposition, armed with a database of citizens' political allegiances, addresses and photos. Yesterday's despot would have

needed an army of soldiers to accomplish this task, and those soldiers could be fooled, bribed, or made to lose their cool and shoot the wrong people.

The despots of tomorrow will just buy a few thousand automated gun drones. Thanks to Moore's Law, which describes the exponential increase in computing power per dollar since the invention of the transistor, the price of a drone with reasonable AI will one day become as accessible as an AK-47. Three or four sympathetic software engineers can reprogram the drones to patrol near the dissidents' houses and workplaces and shoot them on sight. The drones would make fewer mistakes, they wouldn't be swayed by bribes or sob stories, and above all, they'd work much more efficiently than human soldiers, allowing the ambitious despot to mop up the detractors before the international community can marshall a response.

Repurposing an AI machine for destructive purposes will be far easier than repurposing a nuclear reactor.

Because of the massive increase in efficiency brought about by automation, AI weapons will lower the barrier to entry for deranged individuals looking to perpetrate such atrocities. What was once the sole domain of dictators in control of an entire army will be brought within reach of moderately wealthy individuals.

Manufacturers and governments interested in developing such weapons may claim that they can engineer proper safeguards to ensure that they cannot be reprogrammed or hacked. Such claims should be greeted with skepticism. Electronic voting machines, ATMs, blu-ray disc players, and even cars speeding down the highway have all been recently compromised in spite of their advertised security. History demonstrates that a computing device tends to eventually yield to a motivated hacker's attempts to repurpose it. AI weapons are unlikely to be an exception.

International treaties going back to 1925 have banned the use of chemical and biological weapons in warfare. The use of hollow-point bullets was banned even earlier, in 1899. The reasoning is that such weapons create extreme and unnecessary suffering. They are especially prone to civilian casualties, such as when people inhale poison gas, or when doctors are injured in attempting to remove a hollow-point bullet. All of these weapons are prone to generate indiscriminate suffering and death, and so they are banned.

Is there a class of AI machines that is equally worthy of a ban? The answer, unequivocally, is yes. If an AI machine can be cheaply and easily converted into an effective and indiscriminate mass killing device, then there should be an international convention against it. Such machines are not unlike radioactive metals. They can be used for reasonable purposes. But we must carefully control them because they can be easily converted into devastating weapons. The difference is that repurposing an AI machine for destructive purposes will be far easier than repurposing a nuclear reactor.

What counts as intelligence, and what counts as a weapon?

We should ban AI weapons not because they are all immoral. We should ban them because humans will transform AI weapons into hideous blood-thirsty monsters using mods and hacks easily found online. A simple piece of code will transform many AI weapons into killing machines capable of the worst excesses of chemical weapons, biological weapons, and hollow-point bullets.

Banning certain kinds of artificial intelligence requires grappling with a number of philosophical questions. Would an AI weapons ban have prohibited the US Strategic Defense Initiative, popularly known as the Star Wars missile defense? Cars can be used as weapons, so does the petition propose to ban Google's self-driving cars, or the self-driving cars being deployed in cities around the UK? What counts as intelligence, and what counts as a weapon?

These are difficult and important questions. However, they do not need to be answered before we agree to formulate a convention to control AI weapons. The limits of what's acceptable must be seriously considered by the international community, and through the advice of scientists, philosophers, and computer engineers. The U.S. Department of Defense already prohibits fully autonomous weapons in some sense. It is time to refine and expand that prohibition to an international level.

Of course, no international ban will completely stop the spread of AI weapons. But this is no reason to scrap the ban. If we as a community think there is reason to ban chemical weapons, biological weapons, and hollow-point bullets, then there is reason to ban AI weapons too.

Our Fear of Artificial Intelligence

Paul Ford, *The MIT Technology Review,* February 11, 2015.
https://www.technologyreview.com/s/534871/our-fear-of-artificial-intelligence/

Editors' Note:
This article summarizes some of the possible "end of the world" scenarios, where AI becomes a sort of super intelligence that humanity is unable to stop. It summarizes some of the work of Nick Bostrom, an important critic of current AI development, who discusses possible unforeseen developments that would cause AI systems to intentionally or unintentionally bring about world ending scenarios. Many experts disagree with these predictions and their rebuttals are included in the article as well.
~

Years ago I had coffee with a friend who ran a startup. He had just turned 40. His father was ill, his back was sore, and he found himself overwhelmed by life. "Don't laugh at me," he said, "but I was counting on the singularity."

My friend worked in technology; he'd seen the changes that faster microprocessors and networks had wrought. It wasn't that much of a step for him to believe that before he was beset by middle age, the intelligence of machines would exceed that of humans—a moment that futurists call the singularity. A benevolent superintelligence might analyze the human genetic code at great speed and unlock the secret to eternal youth. At the very least, it might know how to fix your back.

But what if it wasn't so benevolent? Nick Bostrom, a philosopher who directs the Future of Humanity Institute at the University of Oxford, describes the following scenario in his book Superintelligence, which has prompted a great deal of debate about the future of artificial intelligence. Imagine a machine that we might call a "paper-clip maximizer"—that is, a machine programmed to make as many paper clips as possible. Now imagine that this machine somehow became incredibly intelligent. Given its goals, it might then decide to create new, more efficient paper-clip-manufacturing machines—until, King Midas style, it had converted essentially everything to paper clips.

No worries, you might say: you could just program it to make exactly a million paper clips and halt. But what if it makes the paper clips and then decides to check its work? Has it counted correctly? It needs to become smarter to be sure. The superintelligent machine manufactures some as-yet-uninvented raw-computing material (call it "computronium") and uses that to check each doubt. But each new doubt yields further digital doubts, and so on, until the entire earth is converted to computronium. Except for the million paper clips.

Bostrom does not believe that the paper-clip maximizer will come to be, exactly; it's a thought experiment, one designed to show how even careful system design can fail to restrain extreme machine intelligence. But he does believe that superintelligence could emerge, and while it could be great, he thinks it could also decide it doesn't need humans around. Or do any number of other things that destroy the world. The title of chapter 8 is: "Is the default outcome doom?"

If this sounds absurd to you, you're not alone. Critics such as the robotics pioneer Rodney Brooks say that people who fear a runaway AI misunderstand what computers are doing when we say they're thinking or getting smart. From this perspective, the putative superintelligence Bostrom describes is far in the future and perhaps impossible.

Yet a lot of smart, thoughtful people agree with Bostrom and are worried now. Why?

Volition

The question "Can a machine think?" has shadowed computer science from its beginnings. Alan Turing proposed in 1950 that a machine could be taught like a child; John McCarthy, inventor of the programming language LISP, coined the term "artificial intelligence" in 1955. As AI researchers in the 1960s and 1970s began to use computers to recognize images, translate between languages, and understand instructions in normal language and not just code, the idea that computers would eventually develop the ability to speak and think—and thus to do evil—bubbled into mainstream culture. Even beyond the oft-referenced HAL from 2001: A Space Odyssey, the 1970 movie Colossus: The Forbin Project featured a large blinking mainframe computer that brings the world to the brink of nuclear destruction; a similar theme was explored 13 years later in WarGames. The androids of 1973's Westworld went crazy and started killing. Extreme AI predictions are "comparable to seeing more efficient internal combustion engines… and jumping to the conclusion that the warp drives are just around the corner," Rodney Brooks writes.

When AI research fell far short of its lofty goals, funding dried up to a trickle, beginning long "AI winters." Even so, the torch of the intelligent machine was carried forth in the 1980s and '90s by sci-fi authors like Vernor Vinge, who popularized the concept of the singularity; researchers like the roboticist Hans Moravec, an expert in computer vision; and the engineer/entrepreneur Ray Kurzweil, author of the 1999 book The Age of Spiritual Machines. Whereas Turing had posited a humanlike intelligence, Vinge, Moravec, and Kurzweil were thinking bigger: when a computer became capable of independently devising ways to achieve goals, it would very likely be capable of introspection—and thus able to modify its software and make itself more intelligent. In short order, such a computer would be able to design its own hardware.

As Kurzweil described it, this would begin a beautiful new era. Such machines would have the insight and patience (measured in picoseconds) to solve the outstanding problems of nanotechnology and spaceflight; they would improve the human condition and let us upload our consciousness into an immortal digital form. Intelligence would spread throughout the cosmos.

You can also find the exact opposite of such sunny optimism. Stephen Hawking has warned that because people would be unable to compete with an advanced AI, it "could spell the end of the human race." Upon reading Superintelligence, the entrepreneur Elon Musk tweeted: "Hope we're not just the biological boot loader for digital superintelligence. Unfortunately, that is increasingly probable." Musk then followed with a $10 million grant to the Future of Life

Institute. Not to be confused with Bostrom's center, this is an organization that says it is "working to mitigate existential risks facing humanity," the ones that could arise "from the development of human-level artificial intelligence."

No one is suggesting that anything like superintelligence exists now. In fact, we still have nothing approaching a general-purpose artificial intelligence or even a clear path to how it could be achieved. Recent advances in AI, from automated assistants such as Apple's Siri to Google's driverless cars, also reveal the technology's severe limitations; both can be thrown off by situations that they haven't encountered before. Artificial neural networks can learn for themselves to recognize cats in photos. But they must be shown hundreds of thousands of examples and still end up much less accurate at spotting cats than a child.

This is where skeptics such as Brooks, a founder of iRobot and Rethink Robotics, come in. Even if it's impressive—relative to what earlier computers could manage—for a computer to recognize a picture of a cat, the machine has no volition, no sense of what cat-ness is or what else is happening in the picture, and none of the countless other insights that humans have. In this view, AI could possibly lead to intelligent machines, but it would take much more work than people like Bostrom imagine. And even if it could happen, intelligence will not necessarily lead to sentience. Extrapolating from the state of AI today to suggest that superintelligence is looming is "comparable to seeing more efficient internal combustion engines appearing and jumping to the conclusion that warp drives are just around the corner," Brooks wrote recently on Edge.org. "Malevolent AI" is nothing to worry about, he says, for a few hundred years at least.

Insurance policy

Even if the odds of a superintelligence arising are very long, perhaps it's irresponsible to take the chance. One person who shares Bostrom's concerns is Stuart J. Russell, a professor of computer science at the University of California, Berkeley. Russell is the author, with Peter Norvig (a peer of Kurzweil's at Google), of Artificial Intelligence: A Modern Approach, which has been the standard AI textbook for two decades.

"There are a lot of supposedly smart public intellectuals who just haven't a clue," Russell told me. He pointed out that AI has advanced tremendously in the last decade, and that while the public might understand progress in terms of Moore's Law (faster computers are doing more), in fact recent AI work has been fundamental, with techniques like deep learning laying the groundwork for computers that can automatically increase their understanding of the world around them.

[…]

Because Google, Facebook, and other companies are actively looking to create an intelligent, "learning" machine, he reasons, "I would say that one of the things we ought not to do is to press full steam ahead on building superintelligence without giving thought to the potential risks. It just seems a bit daft." Russell made an analogy: "It's like fusion research. If you ask a fusion

researcher what they do, they say they work on containment. If you want unlimited energy you'd better contain the fusion reaction." Similarly, he says, if you want unlimited intelligence, you'd better figure out how to align computers with human needs.

Bostrom's book is a research proposal for doing so. A superintelligence would be godlike, but would it be animated by wrath or by love? It's up to us (that is, the engineers). Like any parent, we must give our child a set of values. And not just any values, but those that are in the best interest of humanity. We're basically telling a god how we'd like to be treated. How to proceed?

Bostrom draws heavily on an idea from a thinker named Eliezer Yudkowsky, who talks about "coherent extrapolated volition"—the consensus-derived "best self" of all people. AI would, we hope, wish to give us rich, happy, fulfilling lives: fix our sore backs and show us how to get to Mars. And since humans will never fully agree on anything, we'll sometimes need it to decide for us—to make the best decisions for humanity as a whole. How, then, do we program those values into our (potential) superintelligences? What sort of mathematics can define them? These are the problems, Bostrom believes, that researchers should be solving now. Bostrom says it is "the essential task of our age."

For the civilian, there's no reason to lose sleep over scary robots. We have no technology that is remotely close to superintelligence. Then again, many of the largest corporations in the world are deeply invested in making their computers more intelligent; a true AI would give any one of these companies an unbelievable advantage. They also should be attuned to its potential downsides and figuring out how to avoid them.

This somewhat more nuanced suggestion—without any claims of a looming AI-mageddon—is the basis of an open letter on the website of the Future of Life Institute, the group that got Musk's donation. Rather than warning of existential disaster, the letter calls for more research into reaping the benefits of AI "while avoiding potential pitfalls." This letter is signed not just by AI outsiders such as Hawking, Musk, and Bostrom but also by prominent computer scientists (including Demis Hassabis, a top AI researcher). You can see where they're coming from. After all, if they develop an artificial intelligence that doesn't share the best human values, it will mean they weren't smart enough to control their own creations.