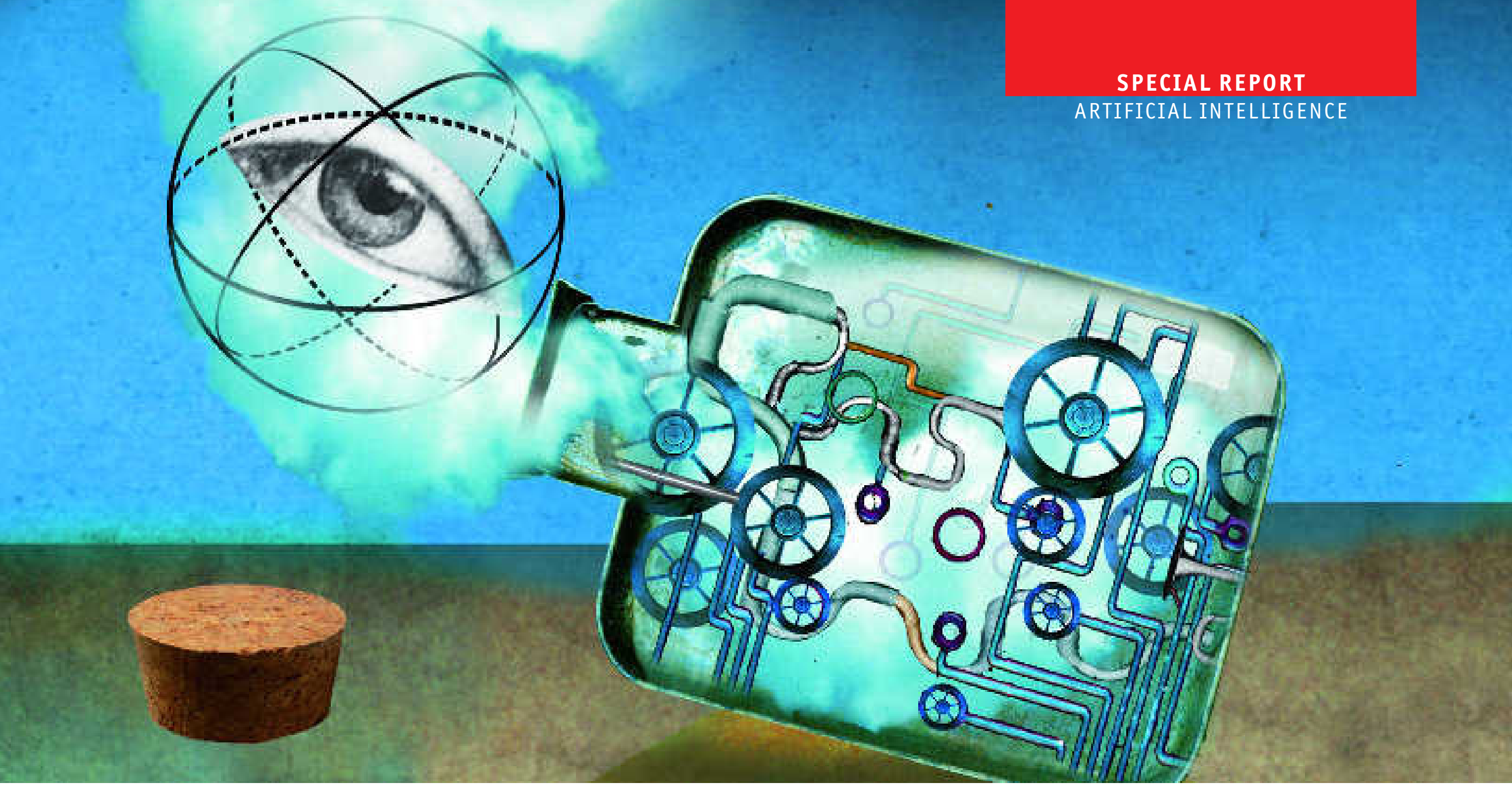
The Economist

SPECIAL REPORT
ARTIFICIAL INTELLIGENCE
June 25th 2016

The return of the machinery question



The return of the machinery question

After many false starts, artificial intelligence has taken off. Will it cause mass unemployment or even destroy mankind? History can provide some helpful clues, says Tom Standage

THERE IS SOMETHING familiar about fears that new machines will take everyone's jobs, benefiting only a select few and upending society. Such concerns sparked furious arguments two centuries ago as industrialisation took hold in Britain. People at the time did not talk of an "industrial revolution" but of the "machinery question". First posed by the economist David Ricardo in 1821, it concerned the "influence of machinery on the interests of the different classes of society", and in particular the "opinion entertained by the labouring class, that the employment of machinery is frequently detrimental to their interests". Thomas Carlyle, writing in 1839, railed against the "demon of mechanism" whose disruptive power was guilty of "oversetting whole multitudes of workmen".

Today the machinery question is back with a vengeance, in a new guise. Technologists, economists and philosophers are now debating the implications of artificial intelligence (AI), a fast-moving technology that enables machines to perform tasks that could previously be done only by humans. Its impact could be profound. It threatens workers whose jobs had seemed impossible to automate, from radiologists to legal clerks. A widely cited study by Carl Benedikt Frey and Michael Osborne of Oxford University, published in 2013, found that 47% of jobs in America were at high risk of being "substituted by computer capital" soon. More recently Bank of America Merrill Lynch predicted that by 2025 the "annual creative disruption impact" from AI could amount to $14 trillion-33 trillion, including a $9 trillion reduction in employment costs thanks to AI-enabled automation of knowledge work; cost reductions of $8 trillion in manufacturing and health care; and $2 trillion in efficiency gains from the deployment of self-driving cars and drones. The McKinsey Global Institute, a think-tank, says AI is contributing to a transformation of society "happening ten times faster and at 300 times the scale, or roughly 3,000 times the impact" of the Industrial Revolution.

CONTENTS

A list of sources is at
Economist.com/specialreports

ACKNOWLEDGMENTS

In addition to those quoted in this report, the author wishes to thank the following people for their assistance and guidance during its preparation: Itamar Arel, Azeem Azhar, Donna Dubinsky, Jeff Hawkins, Puneet Mehta, Ali Parsa, Murray Shanahan, Ivana Schnur, Richard Susskind and Mike Tung.

Just as people did two centuries ago, many fear that machines will make millions of workers redundant, causing inequality and unrest. Martin Ford, the author of two bestselling books on the dangers of automation, worries that middle-class jobs will vanish, economic mobility will cease and a wealthy plutocracy could "shut itself away in gated communities or in elite cities, perhaps guarded by autonomous military robots and drones". Others fear that AI poses an existential threat to humanity, because superintelligent computers might not share mankind's goals and could turn on their creators. Such concerns have been expressed, among others, by Stephen Hawking, a physicist, and more surprisingly by Elon Musk, a billionaire technology entrepreneur who founded SpaceX, a rocket company, and Tesla, a maker of electric cars. Echoing Carlyle, Mr Musk warns that "with artificial intelligence, we're summoning the demon." His Tesla cars use the latest AI technology to drive themselves, but Mr Musk frets about a future AI overlord becoming too powerful for humans to control. "It's fine if you've got Marcus Aurelius as the emperor, but not so good if you have Caligula," he says.

It's all Go

Such concerns have been prompted by astonishing recent progress in AI, a field long notorious for its failure to deliver on its promises. "In the past couple of years it's just completely exploded," says Demis Hassabis, the boss and co-founder of DeepMind, an AI startup bought by Google in 2014 for $400m. Earlier this year his firm's AlphaGo system defeated Lee Sedol, one of the world's best players of Go, a board game so complex that computers had not been expected to master it for another decade at least. "I was a sceptic for a long time, but the progress now is real. The results are real. It works," says Marc Andreessen of Andreessen Horowitz, a Silicon Valley venture-capital firm.

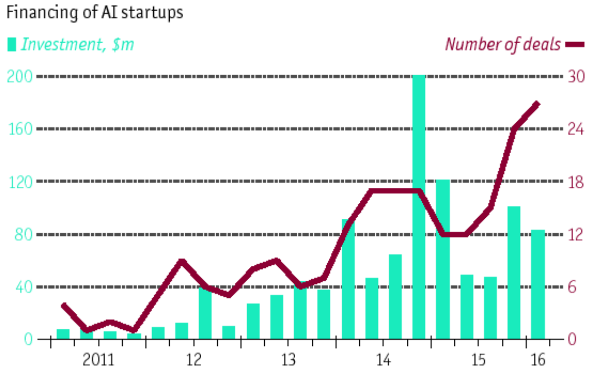
In particular, an AI technique called "deep learning", which allows systems to learn and improve by crunching lots of examples rather than being explicitly programmed, is already being used to power internet search engines, block spam e-mails, suggest e-mail replies, translate web pages, recognise voice commands, detect credit-card fraud and steer self-driving cars. "This is a big deal," says Jen-Hsun Huang, chief executive of NVIDIA, a firm whose chips power many AI systems. "Instead of people writing software, we have data writing software."

Where some see danger, others see opportunity. Investors are piling into the field. Technology giants are buying AI startups and competing to attract the best researchers from academia. In 2015 a record $8.5 billion was spent on AI companies, nearly four times as much as in 2010, according to Quid, a data-analysis company. The number of investment rounds in AI companies in 2015 was 16% up on the year before, when for the technology sector as a whole it declined by 3%, says Nathan Benaich of Playfair Capital, a fund that has 25% of its portfolio invested in AI. "It's the Uber for X" has given way to "It's X plus AI" as the default business model for startups. Google, Facebook, IBM, Amazon and Microsoft are trying to establish ecosystems around AI services provided in the cloud. "This technology will be applied in pretty much every industry out there that has any kind of data—anything from genes to images to language," says Richard Socher, founder of MetaMind, an AI startup recently acquired by Salesforce, a cloud-computing giant. "AI will be everywhere."

What will that mean? This special report will examine the rise of this new technology, explore its potential impact on jobs, education and policy, and consider its ethical and regulatory implications. Along the way it will consider the lessons that can be learned from the original response to the machinery question. AI excites fear and enthusiasm in equal measure, and raises a lot of questions. Yet it is worth remembering that many of those questions have been asked, and answered, before. ■

Hotting up

Financing of AI startups



Source: CB Insights

Technology

From not working to neural networking

The artificial-intelligence boom is based on an old idea, but with a modern twist

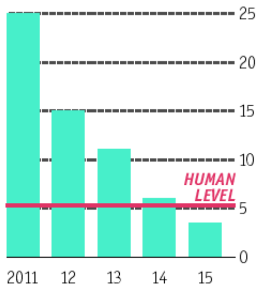
HOW HAS ARTIFICIAL intelligence, associated with hubris and disappointment since its earliest days, suddenly become the hottest field in technology? The term was coined in a research proposal written in 1956 which suggested that significant progress could be made in getting machines to "solve the kinds of problems now reserved for humans…if a carefully selected group of scientists work on it together for a summer". That proved to be wildly overoptimistic, to say the least, and despite occasional bursts of progress, AI became known for promising much more than it could deliver. Researchers mostly ended up avoiding the term, preferring to talk instead about "expert systems" or "neural networks". The rehabilitation of "AI", and the current excitement about the field, can be traced back to 2012 and an online contest called the ImageNet Challenge.

ImageNet is an online database of millions of images, all labelled by hand. For any given word, such as "balloon" or "strawberry", ImageNet contains several hundred images. The annual ImageNet contest encourages those in the field to compete and measure their progress in getting computers to recognise and label images automatically. Their systems are first trained using a set of images where the correct labels are provided, and are then challenged to label previously unseen test images. At a follow-up workshop the winners share and discuss their techniques. In 2010 the winning system could correctly label an image 72% of the time (for humans, the average is 95%). In 2012 one team, led by Geoff Hinton at the University of Toronto, achieved a jump in accuracy to 85%, thanks to a novel technique known as "deep learning". This brought further rapid improvements, producing an accuracy of 96% in the ImageNet Challenge in 2015 and surpassing humans for the first time.

The 2012 results were rightly recognised as a breakthrough, but they relied on "combining pieces that were all there before", says Yoshua Bengio, a computer scientist at the University of Montreal who, along with Mr Hinton and a few others, is recognised as a pioneer of deep learning. In essence, this technique uses huge amounts of computing power and vast quantities of training data to supercharge an old idea from the dawn of AI: so-called artificial neural networks (ANNs). These are biologically inspired networks of artificial neurons, or brain cells.

Ever cleverer

Error rates on ImageNet Visual Recognition Challenge, %



Sources: ImageNet; Stanford Vision Lab

In a biological brain, each neuron can be triggered by other neurons whose outputs feed into it, and its own output can then trigger other neurons in turn. A simple ANN has an input layer of neurons where data can be fed into the network, an output layer where results come out, and possibly a couple of hidden layers in the middle where information is processed. (In practice, ANNs are simulated entirely in software.) Each neuron within the network has a set of "weights" and an "activation function" that controls the firing of its output. Training a neural network involves adjusting the neurons' weights so that a given input produces the desired output (see diagram, next page). ANNs were starting to achieve some useful results in the early 1990s, for example in recognising handwritten numbers. But attempts to get them to do more complex tasks ran into trouble.

In the past decade new techniques and a simple tweak to the activation function has made training deep networks feasible. At the same time the rise of the internet has made billions of documents, images and videos available for training purposes. All this takes a lot of number-crunching power, which became readily available when several AI research groups realised around 2009 that graphical processing units (GPUs), the specialised chips used in PCs and video-games consoles to generate fancy graphics, were also well suited to running deep-learning algorithms. An AI research group at Stanford University led by Andrew Ng, who subsequently moved to Google and now works for Baidu, a Chinese internet giant, found that GPUs could speed up its deep-learning system nearly a hundredfold. Suddenly, training a four-layer neural network, which had previously taken several weeks, took less than a day. It is a pleasing symmetry, says Jen-Hsun Huang, the boss of NVIDIA, which makes GPUs, that the same chips that are used to conjure up imaginary worlds for gamers can also be used to help computers understand the real world through deep learning.

What got people excited about this field is that one technique, deep learning, can be applied to so many different domains

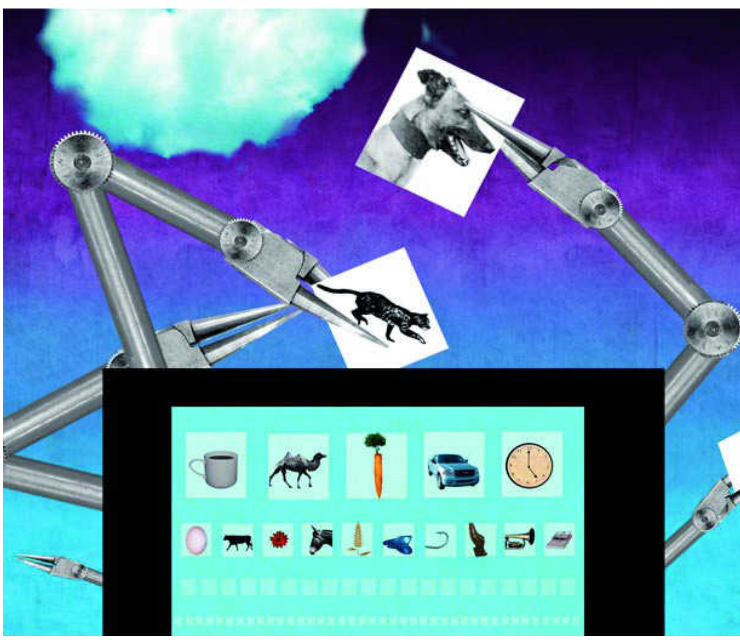
The ImageNet results showed what deep learning could do. Suddenly people started to pay attention, not just within the AI community but across the technology industry as a whole. Deep-learning systems have since become more powerful: networks 20 or 30 layers deep are not uncommon, and researchers at Microsoft have built one with 152 layers. Deeper networks are capable of higher levels of abstraction and produce better results, and these networks have proved to be good at solving a very wide range of problems.

"What got people excited about this field is that one learning technique, deep learning, can be applied to so many different domains," says John Giannandrea, head of machine-intelligence research at Google and now in charge of its search engine too. Google is using deep learning to boost the quality of its web-search results, understand commands spoken into smartphones, help people search their photos for particular images, suggest automatic answers to e-mails, improve its service for translating web pages from one language to another, and help its self-driving cars understand their surroundings.

Learning how to learn

Deep learning comes in many flavours. The most widely used variety is "supervised learning", a technique that can be used to train a system with the aid of a labelled set of examples. For e-mail spam filtering, for example, it is possible to assemble an enormous database of example messages, each of which is labelled "spam" or "not spam". A deep-learning system can be trained using this database, repeatedly working through the examples and adjusting the weights inside the neural network to improve its accuracy in assessing spamminess. The great merit of this approach is that there is no need for a human expert to draw up a list of rules, or for a programmer to implement them in code; the system learns directly from the labelled data.

Systems trained using labelled data are being used to classify images, recognise speech, spot fraudulent credit-card transactions, identify spam and malware, and target advertisements—all applications in which the right answer is known for a large number of previous cases. Facebook can recognise and tag your friends and family when you upload a photograph, and recently launched a system that describes the contents of photographs for blind users ("two people, smiling, sunglasses, outdoor, water"). There is a huge reservoir of data to which supervised learning can be applied, says Mr Ng. Adoption of the technology has allowed existing firms in financial services, computer security and marketing to relabel themselves as AI companies.

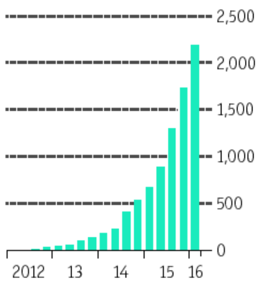


Another technique, unsupervised learning, involves training a network by exposing it to a huge number of examples, but without telling it what to look for. Instead, the network learns to recognise features and cluster similar examples, thus revealing hidden groups, links or patterns within the data.

Unsupervised learning can be used to search for things when you do not know what they look like: for monitoring network traffic patterns for anomalies that might correspond to a cyber-attack, for example, or examining large numbers of insurance claims to detect new kinds of fraud. In a famous example, when working at Google in 2011, Mr Ng led a project called Google Brain in which a giant unsupervised learning system was asked to look for common patterns in thousands of unlabelled YouTube videos. One day one of Mr Ng's PhD students had a surprise for him. "I remember him calling me over to his computer and saying, 'look at this'," Mr Ng recalls. On the screen was a furry face, a pattern distilled from thousands of examples. The system had discovered cats.

Spotting cats

Number of projects at Google using TensorFlow*



*Google's main software library for machine learning

Source: Google

Reinforcement learning sits somewhere in between supervised and unsupervised learning. It involves training a neural network to interact with an environment with only occasional feedback in the form of a reward. In essence, training involves adjusting the network's weights to search for a strategy that consistently generates higher rewards. DeepMind is a specialist in this area. In February 2015 it published a paper in *Nature* describing a reinforcement-learning system capable of learning to play 49 classic Atari video games, using just the on-screen pixels and the game score as inputs, with its output connected to a virtual controller. The system learned to play them all from scratch and achieved human-level performance or better in 29 of them.

Gaming the system

Video games are an ideal training ground for AI research, says Demis Hassabis of DeepMind, because "they are like microcosms of the real world, but are cleaner and more constrained." Gaming engines can also generate large quantities of training data very easily. Mr Hassabis used to work in the video-games industry before taking a PhD in cognitive neuroscience and starting DeepMind. The company now operates as an AI research arm for Google, from offices near King's Cross station in London.

DeepMind made headlines in March when its AlphaGo system defeated Lee Sedol, one of the world's best Go players, by 4-1 in a five-game match in Seoul. AlphaGo is a reinforcement-learning system with some unusual features. It consists of several interconnected modules, including two deep neural networks, each of which specialises in a different thing—just like the modules of the human brain. One of them has been trained by analysing millions of games to suggest a handful of promising moves, which are then evaluated by the other one, guided by a technique that works by random sampling. The system thus combines biologically inspired techniques with non-biologically inspired ones. AI researchers have argued for decades over which approach is superior, but AlphaGo uses both. "It's a hybrid system because we believe we're going to need more than deep learning to solve intelligence," says Mr Hassabis.

He and other researchers are already looking to the next step, called transfer learning. This would allow a reinforcement-learning system to build on previously acquired knowledge, rather than having to be trained from scratch every time. Humans do this effortlessly, notes Mr Hassabis. Mr Giannandrea recalls that his four-year-old daughter was able to tell that a penny-farthing was a kind of bicycle even though she had never seen one before. "Computers can't do that," he says.

MetaMind, an AI startup recently acquired by Salesforce, is pursuing a related approach called multitask learning, where the same neural-network architecture is used to solve several different kinds of problems in such a way that experience of one thing makes it better at another. Like DeepMind, it is exploring modular architectures; one them, called a "dynamic memory network", can, among other things, ingest a series of statements and then answer questions about them, deducing the logical connections between them (Kermit is a frog; frogs are green; so Kermit is green). MetaMind has also combined natural-language and image-recognition networks into a single system that can answer questions about images ("What colour is the car?"). Its technology could be used to power automated customer-service chatbots or call-centres for Salesforce's customers.

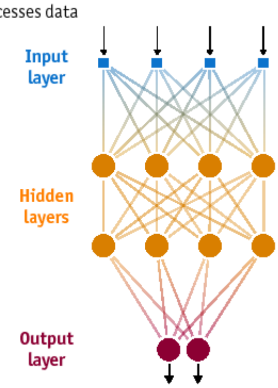
In the past, promising new AI techniques have tended to run out of steam quickly. But deep learning is different. "This stuff actually works," says Richard Socher of MetaMind. People are using it every day without realising it. The long-term goal to which Mr Hassabis, Mr Socher and others aspire is to build an "artificial general intelligence" (AGI)—a system capable of solving a wide range of tasks—rather than building a new AI system for each problem. For years, AI research has focused on solving specific, narrow problems, says Mr Socher, but now researchers are "taking these more advanced Lego pieces and putting them together in new ways". Even the most optimistic of them think it will take another decade to attain human-level AGI. But, says Mr Hassabis, "we think we know what the dozen or so key things are that are required to get us close to something like AGI."

Meanwhile AI is already useful, and will rapidly become more so. Google's Smart Reply system, which uses two neural networks to suggest answers to e-mails, went from being a deep-learning research project to a live product in just four months (though initially it had to be discouraged from suggesting the reply "I love you" to nearly every message). "You can publish a paper in a research journal and literally have a company use that system the next month," says Mr Socher. There is a steady flow of academic papers from AI companies both large and small; AI researchers have been allowed to continue publishing their results in peer-reviewed journals, even after moving into industry. ▸▸

Layer cake

How an artificial neural network processes data

A neural network is organised into layers. Information is fed into the input layer, and artificial neurons in a series of "hidden" layers combine signals by applying different weights to them, and passing the result to the next layer. A "deep" network with many hidden layers can detect increasingly subtle features of the input data. Training the network involves adjusting its internal weights so that it gives the desired response when presented with particular inputs.



Many of them maintain academic posts alongside working for companies. "If you won't let them publish, they won't work for you," explains Chris Dixon of Andreessen Horowitz.

Google, Facebook, Microsoft, IBM, Amazon, Baidu and other firms have also made some of their deep-learning software available free on an open-source basis. In part, this is because their researchers want to publish what they are doing, so it helps with recruitment. A more cynical view is that big internet firms can afford to give away their AI software because they have a huge advantage elsewhere: access to reams of user data for training purposes. This gives them an edge in particular areas, says Shivon Zilis of Bloomberg Beta, an investment fund, but startups are finding ways into specific markets. Drone startups, for example, can use simulation data to learn how to fly in crowded environments. And lots of training data can be found on the internet, says Sam Altman, president of Y Combinator, a startup incubator. He notes that humans can learn from modest amounts of data, which "suggests that intelligence is possible without massive training sets". Startups pursuing less data-hungry approaches to AI include Numenta and Geometric Intelligence.

Pick and mix

Companies are lining up to supply shovels to participants in this AI gold rush. The name that comes up most frequently is NVIDIA, says Mr Dixon; every AI startup seems to be using its GPU chips to train neural networks. GPU capacity can also be rented in the cloud from Amazon and Microsoft. IBM and Google, meanwhile, are devising new chips specifically built to run AI software more quickly and efficiently. And Google, Microsoft and IBM are making AI services such as speech recognition, sentence parsing and image analysis freely available online, allowing startups to combine such building blocks to form new AI products and services. More than 300 companies from a range of industries have already built AI-powered apps using IBM's Watson platform, says Guru Banavar of IBM, doing everything from filtering job candidates to picking wines.

To most people, all this progress in AI will manifest itself as incremental improvements to internet services they already use every day. Search engines will produce more relevant results; recommendations will be more accurate. Within a few years everything will have intelligence embedded in it to some extent, predicts Mr Hassabis. AI technology will allow computer interfaces to become conversational and predictive, not simply driven by menus and icons. And being able to talk to computers will make them accessible to people who cannot read and write, and cannot currently use the internet, says Mr Bengio.

Yet steady improvements can lead to sudden changes when a threshold is reached and machines are able to perform tasks previously limited to humans. Self-driving cars are getting better fast; at some point soon they may be able to replace taxi drivers, at least in controlled environments such as city centres. Delivery drones, both wheeled and airborne, could similarly compete with human couriers. Improved vision systems and robotic technology could allow robots to stack supermarket shelves and move items around in warehouses. And there is plenty of scope for unexpected breakthroughs, says Mr Dixon.

Others are worried, fearing that AI technology could supercharge the existing computerisation and automation of certain tasks, just as steam power, along with new kinds of machinery, seemed poised to make many workers redundant 200 years ago. "Steam has fearfully accelerated a process that was going on already, but too fast," declared Robert Southey, an English poet. He worried that "the discovery of this mighty power" has come "before we knew how to employ it rightly". Many people feel the same way about artificial intelligence today. ■

The impact on jobs

Automation and anxiety

Will smarter machines cause mass unemployment?

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.

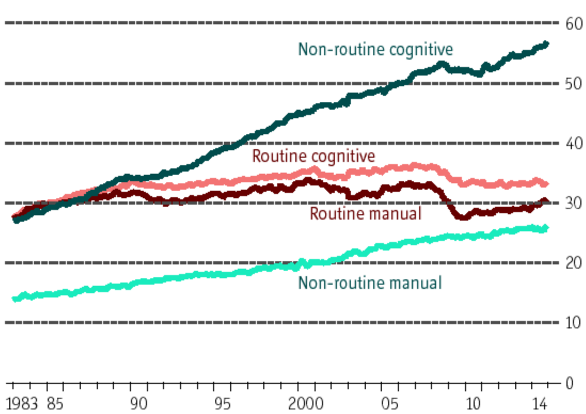
A computer that dispenses expert radiology advice is just one example of how jobs currently done by highly trained white-collar workers can be automated, thanks to the advance of deep learning and other forms of artificial intelligence. The idea that manual work can be carried out by machines is already 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.

Economists are already worrying about "job polarisation", where middle-skill jobs (such as those in manufacturing) are declining but both low-skill and high-skill jobs are expanding. In effect, the workforce bifurcates into two groups doing non-routine work: highly paid, skilled workers (such as architects and senior managers) on the one hand and low-paid, unskilled workers (such as cleaners and burger-flippers) on the other. The stagnation of median wages in many Western countries is cited as evidence that automation is already having an effect—though it is hard to disentangle the impact of offshoring, which has also moved many routine jobs (including manufacturing and call-centre work) to low-wage countries in the developing world. Figures published by the Federal Reserve Bank of St Louis show that in America, employment in non-routine cognitive and non-routine manual jobs has grown steadily since the 1980s, whereas employment in routine jobs has been broadly flat (see chart). As more jobs are automated, this trend seems likely to continue.

Think

United States employment, by type of work, m



Sources: US Population Survey; Federal Reserve Bank of St. Louis

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.

What determines vulnerability to automation is not so much whether the work concerned is manual or white-collar but whether or not it is routine

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. "Never until now did human invention devise such expedients for dispensing with the labour of the poor," said a pamphlet at the time. Subsequent outbreaks of concern occurred in the 1920s ("March of the machine makes idle hands", declared a *New York Times* headline in 1928), the 1930s (when John Maynard Keynes coined the term "technological unemployment") and 1940s, when the *New York Times* referred to the revival of such worries as the renewal of an "old argument".

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.

Catalogue of fears

Probability of computerisation of different occupations, 2013
(1 = certain)

Job	Probability
Recreational therapists	0.003
Dentists	0.004
Athletic trainers	0.007
Clergy	0.008
Chemical engineers	0.02
Editors	0.06
Firefighters	0.17
Actors	0.37
Health technologists	0.40
Economists	0.43
Commercial pilots	0.55
Machinists	0.65
Word processors and typists	0.81
Real-estate sales agents	0.86
Technical writers	0.89
Retail salespeople	0.92
Accountants and auditors	0.94
Telemarketers	0.99

Source: "The Future of Employment: How Susceptible are Jobs to Computerisation?", by C. Frey and M. Osborne (2013)

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.

Only humans need apply

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. ■

Education and policy

Re-educating Rita

Artificial intelligence will have implications for policymakers in education, welfare and geopolitics

IN JULY 2011 Sebastian Thrun, who among other things is a professor at Stanford, posted a short video on YouTube, announcing that he and a colleague, Peter Norvig, were making their "Introduction to Artificial Intelligence" course available free online. By the time the course began in October, 160,000 people in 190 countries had signed up for it. At the same time Andrew Ng, also a Stanford professor, made one of his courses, on machine learning, available free online, for which 100,000 people enrolled. Both courses ran for ten weeks. Mr Thrun's was completed by 23,000 people; Mr Ng's by 13,000.

Such online courses, with short video lectures, discussion boards for students and systems to grade their coursework automatically, became known as Massive Open Online Courses (MOOCs). In 2012 Mr Thrun founded an online-education startup called Udacity, and Mr Ng co-founded another, called Coursera. That same year Harvard University and the Massachusetts Institute of Technology got together to form edX, a non-profit MOOC provider, headed by Anant Agarwal, the head of MIT's artificial-intelligence laboratory. Some thought that MOOCs would replace traditional university teaching. The initial hype around MOOCs has since died down somewhat (though millions of students have taken online courses of some kind). But the MOOC boom illustrated the enormous potential for deliver-

ing education online, in bite-sized chunks.

The fact that Udacity, Coursera and edx all emerged from AI labs highlights the conviction within the AI community that education systems need an overhaul. Mr Thrun says he founded Udacity as an "antidote to the ongoing AI revolution", which will require workers to acquire new skills throughout their careers. Similarly, Mr Ng thinks that given the potential impact of their work on the labour market, AI researchers "have an ethical responsibility to step up and address the problems we cause"; Coursera, he says, is his contribution. Moreover, AI technology has great potential in education. "Adaptive learning"—software that tailors courses for each student individually, presenting concepts in the order he will find easiest to understand and enabling him to work at his own pace—has seemed to be just around the corner for years. But new machine-learning techniques might at last help it deliver on its promise.

Adapt and survive

At the moment, adaptive-learning techniques work best in areas where large numbers of pupils have to learn the same material and a lot of data can be collected, says Mr Ng. Geekie, a Brazilian adaptive-learning startup, guides pupils through the high-school syllabus in thousands of the country's schools. Other startups working in this area include Knewton, Smart Sparrow and DreamBox. Education giants are also paying attention. McGraw-Hill bought ALEKS, another adaptive-learning system, in 2013; Pearson recently announced an expansion of its partnership with Knewton. In a report published in February, Pearson suggests that AI could make learning "more personalised, flexible, inclusive and engaging". Such systems do not replace teachers, but allow them to act as mentors rather than lecturers.

Even outside the AI community, there is a broad consensus that technological progress, and artificial intelligence in particular, will require big changes in the way education is delivered, just as the Industrial Revolution did in the 19th century. As factory jobs overtook agricultural ones, literacy and numeracy became much more important. Employers realised that more educated workers were more productive, but were reluctant to train them themselves because they might defect to another employer. That prompted the introduction of universal state education on a factory model, with schools supplying workers with the right qualifications to work in factories. Industrialisation thus transformed both the need for education and offered a model for providing it. The rise of artificial intelligence could well do the same again, making it necessary to transform educational practices and, with adaptive learning, offering a way of doing so.

"The old system will have to be very seriously revised," says Joel Mokyr of Northwestern University. Since 1945, he points out, educational systems have encouraged specialisation, so students learn more and more about less and less. But as knowledge becomes obsolete more quickly, the most important thing will be learning to relearn, rather than learning how to do one thing very well. Mr Mokyr thinks that education currently treats people too much like clay—"shape it, then bake it, and that's the way it stays"—rather than like putty, which can be reshaped. In future, as more tasks become susceptible to automation, the tasks where human skills are most valuable will constantly shift. "You need to keep learning your entire life—that's been obvious for a long time," says Mr Ng. "What you learn in college isn't enough to keep you going for the next 40 years."

Education will therefore have to be interwoven with full-time work. "People will have to continuously learn new skills to stay current," says Mr Thrun. Hence his firm's focus on "nanodegrees" which can be completed in a few months, alongside a job. Studying for a nanodegree in, say, data science or website programming costs $200 a month, but students who complete a course within 12 months get a 50% refund. A host of websites now offer courses in all kinds of skills, from user-experience design to project management to leadership. Some, like Udacity, charge by the course; others, like Lynda.com, which is owned by LinkedIn, a business-networking site, charge a monthly fee for access to all courses. (It is not difficult to imagine LinkedIn comparing the skill sets of its users against those required to apply for a particular job—and then offering users the courses necessary to fill the gaps.) Users and their potential employers sometimes find it difficult to tell which ones offer good value. More co-operation between government, training providers and employers over certification would help.

America and other developed countries should also put more emphasis on vocational and technical education, as Germany does, rather than encouraging everyone to go to university, says David Autor at MIT. But that does not simply mean offering more apprenticeships, which typically involve five to seven years of training. "That doesn't make sense if the skills you need are changing every three to five years," says James Bessen at the Boston University School of Law. So the traditional apprenticeship model will have to be tweaked. Community colleges are setting up all kinds of schemes that combine education with learning on the job, says Mr Bessen. For example, Siemens, a German industrial giant, has launched a four-year "earn and learn" programme for apprentices at its wind-turbine factory in Charlotte, North Carolina. Apprentices graduate with a degree in mechatronics from a local community college, certification from the local department of labour—and no student debt.

As on-the-job skills come and go, having a solid foundation of basic literacy and numeracy skills will become even more vital. But teaching "soft" skills, too, will be increasingly important. In a paper published in 2013, James Heckman and Tim Kautz of America's National Bureau of Economic Research argue for more emphasis on "character skills" such as perseverance, sociability and curiosity, which are highly valued by employers and correlate closely with employees' ability to adapt to new situations and acquire new skills. Character is a skill, not a trait, they say, and schemes that teach it are both lasting and cost-effective.

Basic attraction

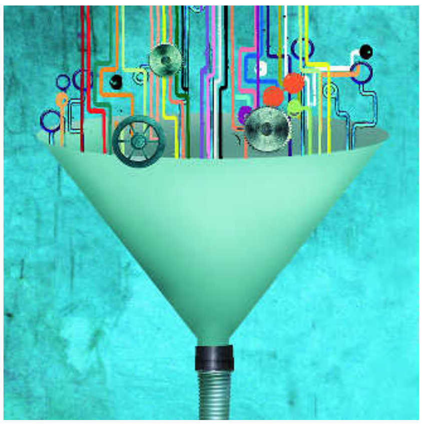
Concerns about AI and automation have also led to calls for a stronger safety net to protect people from labour-market disruption and help them switch to new jobs. In particular, many AI commentators support the idea of a universal basic income: a dramatic simplification of the welfare system that involves paying a fixed amount (say, $10,000 a year) to everyone, regardless of their situation, and doing away with all other welfare payments. Similar ideas were touted during the Industrial Revolution by Thomas Paine and John Stuart Mill, among others. Its chief merit, say its supporters, is that people who are not working, or are working part-time, are not penalised if they decide to work more, because their welfare payments do not decline as their incomes rise. It gives people more freedom to decide how many hours they wish to work, and might also encourage them to retrain by providing them with a small guaranteed income while they do so. Those who predict apocalyptic job destruction see it as a way to keep the consumer economy going and support the non-working population. If most jobs are automated away, an alternative mechanism for redistributing wealth will be needed.

Compared with the complexity of overhauling the education system, a basic income appears to offer a simple, attractive and easily understood solution. The idea enjoys broad support within the technology industry: Y Combinator, a startup incubator, is even funding a study of the idea in Oakland, California. Sam Altman, its president, argues that in a world of rapid technological change, a basic income could help ensure "a smooth transition to the jobs of the future". The idea seems to appeal to techie types in part because of its simplicity and elegance (replacing existing welfare and tax systems, which are like badly written programming code, with a single line) and in part because of its Utopianism. A more cynical view is that it could help stifle complaints about technology causing disruption and inequality, allowing geeks to go on inventing the future unhindered. Mr Altman says that in his experience the techies who support basic income do so for "fairly charitable reasons".

Though it is an attractive idea in principle, the devil is in the details. A universal basic income that replaced existing welfare budgets would be steeply regressive. Divide existing spending on social, pension and welfare schemes (excluding health care) equally, and each citizen would get a basic income of around $6,000 a year in America and $6,200 in Britain, for example (at purchasing-power parity). Compared with existing welfare schemes, that would reduce income for the poorest, while giving the rich money they do not need. But means-testing a basic income risks undermining its simplicity, and thus its low administrative cost. Funding a basic income that would provide a reasonable living would require much higher taxes than at present. Negative income taxes, or schemes such as earned-income tax credits, might be a less elegant but more practical approach.

Many countries, notably Finland and the Netherlands, are planning to experiment with limited forms of basic income next year. A big concern among economists is that a basic income could actually discourage some people from retraining, or indeed working at all—why not play video games all day?—though studies of previous experiments with a basic income suggest that it encourages people to reduce their working hours slightly, rather than giving up work altogether. Another problem is that a basic income is not compatible with open borders and free movement of workers; without restrictions on immigration or entitlement it might attract lots of freeloaders from abroad and cause domestic taxpayers to flee.

This points to another area where policymakers may have to grapple with the impact of advancing automation: its geopolitical implications as it benefits people in some countries more than others. Automation could have a much bigger impact in developing economies than in rich ones, says Mr Autor, because much of what they provide is essentially embodied labour: cheap goods made by low-wage workers, cheap services such as operating call-centres, or doing domestic and construction work overseas. If automation makes rich countries more self-sufficient in these areas, they will have less need for the products and services that have been driving exports and growth in the developing world. Automation could "erode the comparative advantage of much of the developing world", says Mr Autor. Another worry, he says, is that rich countries own the technologies and patents associated with robots and AI, and stand to benefit if they cause a surge in productivity. For the developing world, "it's not clear that they are on the winning side of the bargain" if machines end up outperforming humans in a wide range of activities.



Automation could have a much bigger impact in developing economies than in rich ones

The risk is that automation could deny poorer countries the opportunity for economic development through industrialisation. Economists talk of "premature deindustrialisation"; Dani Rodrik of Harvard University notes that manufacturing employment in Britain peaked at 45% just before the first world war, but has already peaked in Brazil, India and China with a share of no more than 15%. This is because manufacturing is much more automated than it used to be. China recently overtook America as the largest market for industrial automation, according to a report by Citi, a bank, and Oxford University's Martin School. Industrial automation may mean that other emerging economies, such as those in Africa and South America, will find it harder to achieve economic growth by moving workers from fields to factories, and will need to find new growth models. Without manufacturing jobs to build a middle class, observes Tyler Cowen, an economist at George Mason University, such countries "may have high income inequality baked into their core economic structures".

During the Industrial Revolution, John Stuart Mill wrote that "there cannot be a more legitimate object of the legislator's care" than looking after those whose livelihoods are disrupted by machines. At the moment it is mostly rich countries that worry about the effects of automation on education, welfare and development. But policymakers in developing countries will increasingly need to consider them too. ■

Ethics

Frankenstein's paperclips

Techies do not believe that artificial intelligence will run out of control, but there are other ethical worries

AS DOOMSDAY SCENARIOS go, it does not sound terribly frightening. The "paperclip maximiser" is a thought experiment proposed by Nick Bostrom, a philosopher at Oxford University. Imagine an artificial intelligence, he says, which decides to amass as many paperclips as possible. It devotes all its energy to acquiring paperclips, and to improving itself so that it can get paperclips in new ways, while resisting any attempt to divert it from this goal. Eventually it "starts transforming first all of Earth and then increasing portions of space into paperclip manufacturing facilities". This apparently silly scenario is intended to make the serious point that AIs need not have human-like motives or psyches. They might be able to avoid some kinds of human error or bias while making other kinds of mistake, such as fixating on paperclips. And although their goals might seem innocuous to start with, they could prove dangerous if AIs were able to design their own successors and thus repeatedly improve themselves. Even a "fettered superintelligence", running on an isolated computer, might persuade its human handlers to set it free. Advanced AI is not just another technology, Mr Bostrom argues, but poses an existential threat to humanity.

The idea of machines that turn on their creators is not new, going back to Mary Shelley's "Frankenstein" (1818) and earlier; nor is the concept of an AI undergoing an "intelligence explosion" through repeated self-improvement, which was first suggested in 1965. But recent progress in AI has caused renewed concern, and Mr Bostrom has become the best-known proponent of the dangers of advanced AI or, as he prefers to call it, "superintelligence", the title of his bestselling book.

His interest in AI grew out of his analysis of existential threats to humanity. Unlike pandemic disease, an asteroid strike or a supervolcano, the emergence of superintelligence is something that mankind has some control over. Mr Bostrom's book prompted Elon Musk to declare that AI is "potentially more dangerous than nukes". Worries about its safety have also been expressed by Stephen Hawking, a physicist, and Lord Rees, a former head of the Royal Society, Britain's foremost scientific body. All three of them, and many others in the AI community, signed an open letter calling for research to ensure that AI systems are "robust and beneficial"—ie, do not turn evil. Few would disagree that AI needs to be developed in ways that benefit humanity, but agreement on how to go about it is harder to reach.

Mr Musk thinks openness is the key. He was one of the co-founders in December 2015 of OpenAI, a new research institute with more than $1 billion in funding that will carry out AI research and make all its results public. "We think AI is going to have a massive effect on the future of civilisation, and we're trying to take the set of actions that will steer that to a good future," he says. In his view, AI should be as widely distributed as possible. Rogue AIs in science fiction, such as HAL 9000 in "2001: A Space Odyssey" and SKYNET in the "Terminator" films, are big, centralised machines, which is what makes them so dangerous when they turn evil. A more distributed approach will ensure that the benefits of AI are available to everyone, and the consequences less severe if an AI goes bad, Mr Musk argues.

Not everyone agrees with this. Some claim that Mr Musk's real worry is market concentration—a Facebook or Google monopoly in AI, say—though he dismisses such concerns as "petty". For the time being, Google, Facebook and other firms are making much of their AI source code and research freely available in any case. And Mr Bostrom is not sure that making AI technology as widely available as possible is necessarily a good thing. In a recent paper he notes that the existence of multiple AIs "does not guarantee that they will act in the interests of humans or remain under human control", and that proliferation could make the technology harder to control and regulate.

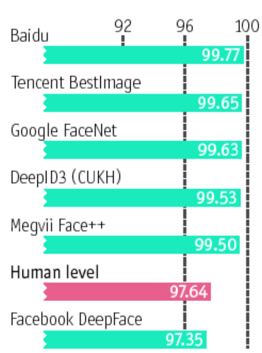
Fears about AIs going rogue are not widely shared by people at the cutting edge of AI research. "A lot of the alarmism comes from people not working directly at the coal face, so they think a lot about more science-fiction scenarios," says Demis Hassabis of DeepMind. "I don't think it's helpful when you use very emotive terms, because it creates hysteria." Mr Hassabis considers the paperclip scenario to be "unrealistic", but thinks Mr Bostrom is right to highlight the question of AI motivation. How to specify the right goals and values for AIs, and ensure they remain stable over time, are interesting research questions, he says. (DeepMind has just published a paper with Mr Bostrom's Future of Humanity Institute about adding "off switches" to AI systems.) A meeting of AI experts held in 2009 in Asilomar, California, also concluded that AI safety was a matter for research, but not immediate concern. The meeting's venue was significant, because biologists met there in 1975 to draw up voluntary guidelines to ensure the safety of recombinant DNA technology.

Sci-fi scenarios

Mr Bostrom responds that several AI researchers do in fact share his concerns, but stresses that he merely wishes to highlight the potential risks posed by AI; he is not claiming that it is dangerous now. For his part, Andrew Ng of Baidu says worrying about superintelligent AIs today "is like worrying about overpopulation on Mars when we have not even set foot on the planet yet", a subtle dig at Mr Musk. (When he is not worrying about AIs, Mr Musk is trying to establish a colony on Mars, as an insurance policy against human life being wiped out on Earth.) AI scares people, says Marc Andreessen, because it combines two deep-seated fears: the Luddite worry that machines will take all the jobs, and the Frankenstein scenario that AIs will "wake up" and do unintended things. Both "keep popping up over and over again". And decades of science fiction have made it a more tangible fear than, say, climate change, which poses a much greater threat.

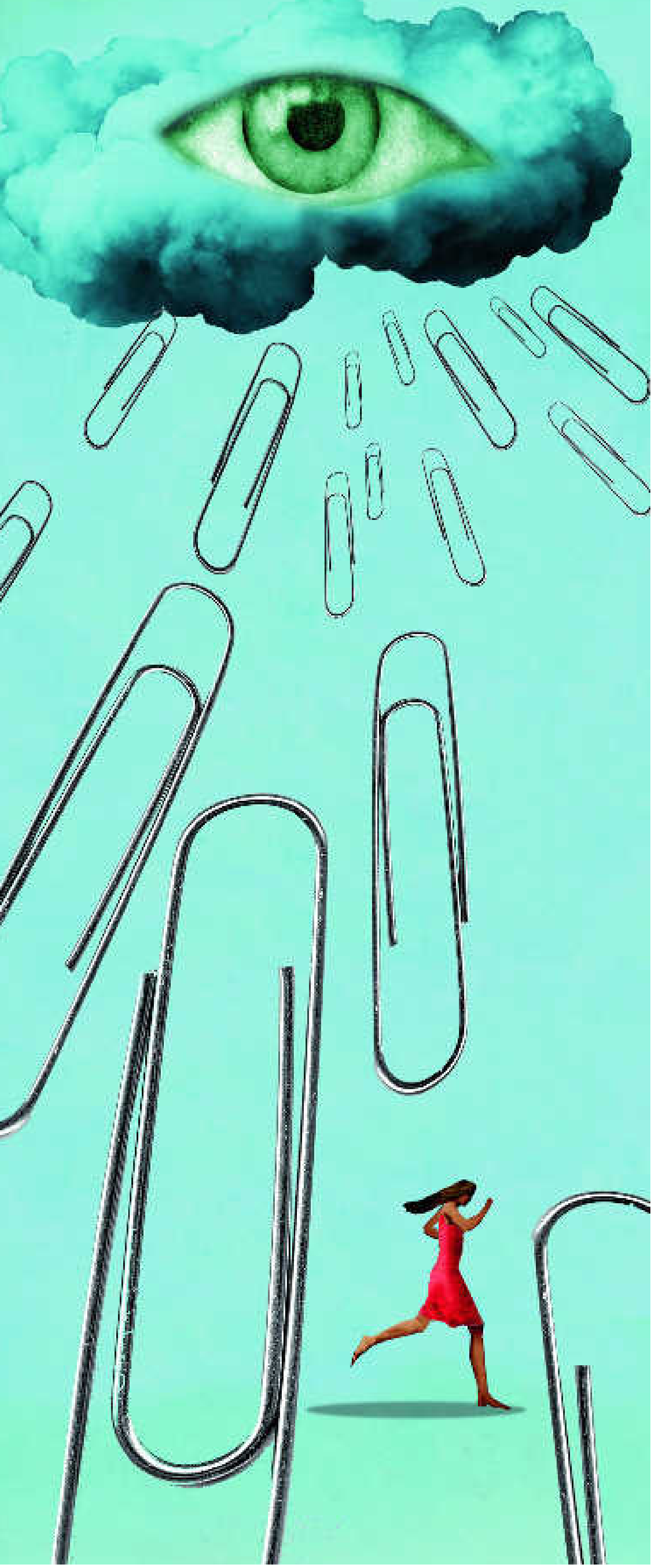
I've just seen a face

Accuracy of facial-recognition algorithms, 2014, %



Source: University of Massachusetts Amherst

AI researchers point to several technical reasons why fear of AI is overblown, at least in its current form. First, intelligence is not the same as sentience or consciousness, says Mr Ng, though all three concepts are commonly elided. The idea that machines will "one day wake up and change their minds about what they will do" is just not realistic, says Francesca Rossi, who works on the ethics of AI at IBM. Second, an "intelligence explosion" is considered unlikely, because it would require an AI to make each version of itself in less ▸▸



time than the previous version as its intelligence grows. Yet most computing problems, even much simpler ones than designing an AI, take much longer as you scale them up.

Third, although machines can learn from their past experiences or environments, they are not learning all the time. A self-driving car, for example, is not constantly retraining itself on each journey. Instead, deep-learning systems have a training phase in which neural-network parameters are adjusted to build a computational model that can perform a particular task, a number-crunching process that may take several days. The resulting model is then deployed in a live system, where it can run using much less computing horsepower, allowing deep-learning models to be used in cars, drones, apps and other products. But those cars, drones and so on do not learn in the wild. Instead, the data they gather while out on a mission are sent back and used to improve the model, which then has to be redeployed. So an individual system cannot learn bad behaviour in a particular environment and "go rogue", because it is not actually learning at the time.

The black-box problem

Amid worries about rogue AIs, there is a risk that nearer-term ethical and regulatory concerns about AI technologies are being overlooked. Facial-recognition systems based on deep learning could make surveillance systems far more powerful, for example. Google's FaceNet can determine with 99.6% accuracy whether two pictures show the same person (humans score around 98%). Facebook's DeepFace is almost as good. When the social-network giant recently launched an app called Moments, which automatically gathers together photos of the same person, it had to disable some of its facial-recognition features in Europe to avoid violating Irish privacy laws.

In Russia, meanwhile, there has been a recent outcry over an app called FindFace, which lets users take photos of strangers and then determines their identity from profile pictures on social networks. The app's creators say it is merely a way to make contact with people glimpsed on the street or in a bar. Russian police have started using it to identify suspects and witnesses. The risk is clear: the end of public anonymity. Gigapixel images of a large crowd, taken from hundreds of metres away, can be analysed to find out who went on a march or protest, even years later. In effect, deep learning has made it impossible to attend a public gathering without leaving a record, unless you are prepared to wear a mask. (A Japanese firm has just started selling Privacy Visor, a funny-looking set of goggles designed to thwart facial-recognition systems.)

Deep learning, with its ability to spot patterns and find clusters of similar examples, has obvious potential to fight crime—and allow authoritarian governments to spy on their citizens. Chinese authorities are analysing people's social-media profiles to assess who might be a dissident, says Patrick Lin, a specialist in the ethics of AI at Stanford Law School. In America, meanwhile, police in Fresno, California, have been testing a system called "Beware" that works out how dangerous a suspect is likely to be, based on an analysis of police files, property records and social-media posts. Another system, called COMPAS, provides guidance when sentencing criminals, by predicting how likely they are to reoffend. Such systems, which are sure to be powered by deep learning soon if they are not already, challenge "basic notions about due process", says Mr Lin.

A related concern is that as machine-learning systems are embedded into more and more business processes, they could be unwittingly discriminatory against particular groups of people. In one infamous example, Google had to apologise when the automatic tagging system in its Photos app labelled black people as "gorillas". COMPAS has been accused of discriminating against black people. AI technology "is already touching people's lives, so it's important that it does not incorporate biases", says Richard Socher of MetaMind. Nobody sets out to make a system racist, he says, but "if it trains on terrible data it will make terrible predictions." Increasingly it is not just intellectual work, but also moral thinking and decision-making, says Mr Lin, that is being done "by what are in effect black boxes".

Fortunately there are ways to look inside these black boxes and determine how they reach their conclusions. An image-processing neural network, for example, can be made to highlight the regions of an input image which most influenced its decision. And many researchers are working on varieties of a technique called "rule extraction" which allows neural networks to explain their reasoning, in effect. The field in which this problem has received most attention is undoubtedly that of self-driving cars.

Such vehicles raise other ethical issues, too, particularly when it comes to how they should behave in emergencies. For example, should a self-driving car risk injuring its occupants to avoid hitting a child who steps out in front of it? Such questions are no longer theoretical. Issues such as who is responsible in an accident, how much testing is required and how to set standards need to be discussed now, says Mr Hassabis. Mr Ng comes at the question from a different angle, suggesting that AI researchers have a moral imperative to build self-driving cars as quickly as possible in order to save lives: most of the 3,000 people who die in car accidents every day are victims of driver error. But even if self-driving cars are much safer, says Daniel Susskind, an economist at Oxford University, attitudes will have to change. People seem to tolerate road deaths caused by humans, but hold machines to much higher standards. "We compare machines to perfection, not to humans doing the same tasks," he says.

Killer app

Many people are worried about the military use of AI, in particular in autonomous weapons that make life-and-death decisions without human intervention. Yoshua Bengio of the University of Montreal says he would like an "outright ban" on the military use of AI. Life-and-death decisions should be made by humans, he says, not machines—not least because machines cannot be held to account afterwards. Mr Hassabis agrees. When Google acquired his firm, he insisted on a guarantee that its technology would not be used for military purposes. He and Mr Bengio have both signed an open letter calling for a ban on "offensive autonomous weapons". (Ronald Arkin of the Georgia Institute of Technology, by contrast, argues that AI-powered military robots might in fact be ethically superior to human soldiers; they would not rape, pillage or make poor judgments under stress.)

Another of Mr Hassabis's ideas, since borrowed by other AI firms, was to establish an ethics board at DeepMind, including some independent observers (though the company has been criticised for refusing to name the board's members). Even if AI firms disagree with the alarmists, it makes sense for them to demonstrate that there are at least some things they think are worth worrying about, and to get involved in regulation before it is imposed from outside. But AI seems unlikely to end up with its own regulatory agency on the lines of America's Federal Aviation Authority or Food and Drug Administration, because it can be applied to so many fields. It seems most likely that AI will require existing laws to be updated, rather than entirely new laws to be passed. The most famous rules governing the behaviour of AI systems are of course the "Three Laws of Robotics" from Isaac Asimov's robot stories. What made the stories interesting was that the robots went wrong in unexpected ways, because the laws simply do not work in practice. It will soon be time to agree on laws that do. ■

Conclusion

Answering the machinery question

Glimpses of an AI-enabled future

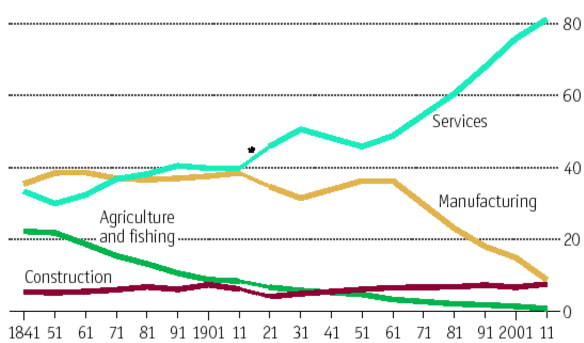
THE ORIGINAL MACHINERY question, which had seemed so vital and urgent, eventually resolved itself. Despite the fears expressed by David Ricardo, among others, that "substitution of machinery for human labour…may render the population redundant", the overall effect of mechanisation turned out to be job creation on an unprecedented scale. Machines allowed individual workers to produce more, reducing the price of many goods, increasing demand and generating a need for more workers. Entirely new jobs were created to oversee the machines. As companies got bigger, they required managers, accountants and other support staff. And whole new and hitherto unimagined industries sprang up with the arrival of the railways, telegraphy and electrification.

To be sure, all this took time. Industrialisation caused pervasive labour-market upheaval as some jobs vanished, others changed beyond recognition and totally new ones emerged. Conditions in factories were grim, and it took several decades before economic growth was reflected in significant wage gains for workers—a delay known as "Engels' pause".

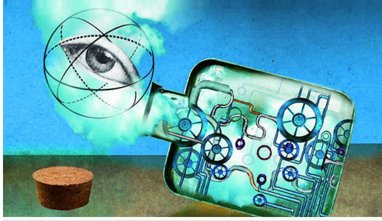
Worries about unemployment gave way to a much wider argument about employment conditions, fuelling the rise of socialist and communist ideas and creating the modern labour movement. By the end of the 19th century the machinery question had faded away, because the answer was so obvious. In 1896 Arthur Hadley, an American economist, articulated the view of the time when he observed that rather than destroying jobs, mechanisation had brought about "a conspicuous increase of employment in those lines where improvements in machinery have been greatest".

What does all this tell us today? Historical analogies are never perfect, but they can be informative. Artificial intelligence is now prompting many of the same concerns as mechanisation did two centuries ago. The 19th-century experience of industrialisation suggests that jobs will be redefined, rather than destroyed; that new industries will emerge; that work and leisure

Jobs come and go
Share of employment in Britain by industry, %



Source: ONS
*England and Wales after 1911



will be transformed; that education and welfare systems will have to change; and that there will be geopolitical and regulatory consequences.

In many ways, the two big debates about AI—whether it will destroy jobs, and whether it might destroy humanity—are really arguments about the rate of change. If you believe that AI is improving so rapidly that human-level artificial general intelligence (AGI) is just around the corner, you are more likely to worry about unexpected and widespread job losses and the possibility that the technology may suddenly get out of control. It seems more probable, however, that AI will improve steadily, and that its impact over the next decade or two, while significant, will not be on the same scale as the epochal shift from a mostly agricultural to a mostly industrial economy.

The debates about whether AI will destroy jobs, and whether it might destroy humanity, are really arguments about the rate of change

AGI is probably still a couple of decades away, perhaps more, so the debate about what it might or might not be able to do, and how society should respond to it, is still entirely theoretical. This special report has therefore focused on the practical effects of AI in the nearer term. These are likely to be a broadening and quickening of the spread of computers into the workplace and everyday life, requiring people to update their skills faster and more frequently than they do at the moment. Provided educational systems are upgraded and made more flexible, which is beginning to happen, that should be entirely feasible.

So far the debate has been dominated by the gloomy possibilities of massive job losses and rogue AIs. More positive scenarios, in which AI dramatically changes the world for the better, tend to attract less attention. So here are three examples. First, AI could transform transport and urban life, starting with self-driving vehicles. Being able to summon one at will could remove the need to own a car, greatly reduce the number of vehicles on the roads and all but eliminate road deaths. Urban environments will enjoy a renaissance as pollution declines and space previously devoted to parking is reallocated to parks, housing and bicycle paths.

Second, AI could soon enable people to converse with a wide range of things: their home and their car, most obviously, just as people talk to a disembodied computer in "Star Trek", but also AI avatars of companies and other organisations, information services, AI advisers and tutors. A host of AI-powered personal assistants, such as Alexa, Cortana, Siri and Viv, are already jostling for position, and could become an important new way to interact with computers and access information, like the web browser and touchscreen before them. Speech alone is not always the best way to interact with a computer, so such conversations will often be accompanied by graphics (perhaps in the form of "augmented reality" overlays on people's vision). AI also has huge potential to help humans talk to one another, by facilitating real-time translation between people using different languages. Basic versions of this technology exist today, and will get better.

The indefatigable helper

Third, AI could make a big difference by turbocharging scientific and medical research. "The thing that excites me the most is using AI to help speed up scientific breakthroughs," says Demis Hassabis of DeepMind. An AI could act as a relentless research assistant, he reckons, in fields from cancer research to climate change, helping solve problems by sifting through data, reading thousands of scientific papers and suggesting hypotheses or pointing out correlations that might be worth investigating. IBM is already working in this area, using its Watson AI technology to analyse large volumes of medical data. Deep learning will be used to analyse the data from the "100,000 Genomes" project now under way in England's National Health Service; the same techniques can help physicists sift reams of data from particle colliders for new discoveries.

After years of frustration with AI's slow rate of progress, it is ironic that many now think it is moving too quickly. Yet a sober assessment suggests that AI should be welcomed, not feared. In the 1840s John Stuart Mill wrote that "the proof of the ultimate benefit to labourers of mechanical inventions…will hereafter be seen to be conclusive." A future economist may say the same of the benefits of AI, not just for labourers but for everyone. ■